

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 07:13:37 ; Search time 486 Seconds

(without alignments)

10560.297 Million cell updates/sec

Title: US-10-024-197-24

Perfect score: 2279

Sequence: 1 agctaaagcaggtacctgca.....aaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC

Gapex 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1659.4	72.8	1661	20	Complete coding se
2	1659.4	72.8	1661	21	Human lysosomal gl
3	1654.2	72.6	1741	22	Unmodified gcc DNA
4	1654.2	72.6	1741	22	Modified gcc DNA
5	1649.8	72.4	1661	9	Encodes recombinan
6	1601.6	70.3	1642	18	Human glucocerebro
7	1587	69.6	1611	22	Human glucocerebro
8	1547.8	67.9	1551	22	Human glucocerebro
9	1544	67.7	1593	24	Nucleotide sequenc

10	1543	67.7	1551	22	AA510863	Human cDNA encodin
11	1543	67.7	1551	24	AAK40235	DNA encoding human
12	1529.6	67.1	1592	24	AAH99924	Nucleotide sequenc
13	667	29.3	7620	14	AAQ39286	Glucocerebrosidase
14	639.4	28.1	5769	24	ABL67250	Thyroid cancer rel
15	363	15.9	417	21	AAAC00312	Human secreted pro
16	209.6	9.2	229	16	AAAT00128	Human gene signatu
17	191.2	8.4	834	22	AAAL20417	Human breast cance
18	175.2	7.7	384	23	ABV37314	Human prostate exp
19	134.8	5.9	246	16	AAAT5845	Human gene signatu
20	130.4	5.7	138	24	ABL38256	Human colon tumour
21	106	4.7	243	23	ABV07380	Human prostate exp
22	92.4	4.1	4861	23	ABL16351	Drosophila melanog
23	89.4	3.9	358	14	AAQ48243	IYS2+1 mutant gluc
24	78.6	3.4	9517	23	ABL16350	Drosophila melanog
25	77	3.4	89	14	AAQ39304	Mutant glucocerebr
26	75	3.3	75	14	AAQ39287	Wild-type glucocer
27	68	3.0	412	22	AAAF67412	Novel human polynu
28	60	2.6	60	24	ABN32368	Human spliced tran
29	59.6	2.6	159	22	AAAL1519	Human breast cance
30	55	2.3	1528	22	AAAF6268	Trichoderma harzia
31	47.4	2.1	65	24	ABN53669	Mouse spliced tran
32	45.4	2.0	55	24	AAH99930	Oligonucleotide 2
33	40	1.8	40	22	AAAF28290	Unmodified gcc con
34	39.4	1.7	41	22	AAAF28288	Human ovarian can
35	38.4	1.7	477	24	ABL80290	Human full-length
36	38.4	1.7	2367	22	AAK94257	Human encoding novel
37	38.4	1.7	2442	22	AAK94333	Human full-length
38	38.4	1.7	2673	21	AAAF6488	Human ORFX ORF2043
39	38.4	1.7	3919	24	ABK54223	Human HEAT-3 DNA
40	38.4	1.7	3950	24	AAAD27270	Human transporter
41	38.4	1.7	3950	24	AAAF27270	DNA encoding novel
42	38.4	1.7	9561	23	AAAF27270	Human secreted pro
43	38.2	1.7	390	21	AAAC04310	Human colon cancer
44	37.8	1.7	399	22	AAH34387	Human reproductive
45	37.6	1.6	19286	22	AAAL03002	

ALIGNMENTS

RESULT 1

AA26682
ID AAX26682 standard; DNA; 1661 BP.

AC AAX26682;

DT 17-JUN-1999 (first entry)

DE Complete coding sequence for lysosomal glucocerebrosidase.

DE Human lysosomal glucocerebrosidase; glycosylated; Gaucher's disease; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 59..1609

FT /*tag= a

FT /*product= "lysosomal glucocerebrosidase"

PW US5879680-A.

XX 09-MAR-1999.

XX 26-MAY-1995;

XX 95US-0452398.

XX 23-DEC-1987;

XX 87US-0137796.

PR 05-FEB-1990;

PR 90US-0474307.

PR 06-AUG-1992;

PR 92US-0925333.

PR 13-JAN-1994;

PR 94US-0186256.

XX 26-MAY-1995;

XX 95US-0452398.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX
PI Eliason WK, Ginns EJ, LaMarca ME, Martin B, Maysak KA;
XX WPI; 1999-203905/17.
DR P-PSDB; AAY01595.
XX
XX New compositions containing glucocerebrosidase - comprising
PT glycosylated recombinantly-produced human glucocerebrosidase, used
PT for treating Gaucher's disease
XX
PS Disclosure; Fig 1A-D; 10pp; English.
XX
XX The present sequence encodes a human lysosomal glucocerebrosidase.
CC The specification describes a composition containing glycosylated
CC recombinantly produced human glucocerebrosidase which can be used
CC for treating Gaucher's disease.
XX
SQ Sequence 1661 BP; 374 A; 496 C; 422 G; 369 T; 0 other;

Query Match 72.8%; Score 1659.4; DB 20; Length 1661;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 108 GGAGTTTCAAGTCTCCAGAGAGAAATGCCAAGCCTTTGAGTAGGTAAGCATCAT 167
DB 1 GGAGTTTCAAGTCTCCAGAGAGAAATGCCAAGCCTTTGAGTAGGTAAGCATCAT 60
QY 168 GGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGACAGTGTCTGGGATCAGGTGCCG 227
DB 61 GGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGACAGTGTCTGGGATCAGGTGCCG 120
QY 228 CCCCTGCATCCCTAAAGCTTCGGCTCAGCTCAGCTCGTGTGTGTCTGCAATGCCACATA 287
DB 121 CCCCTGCATCCCTAAAGCTTCGGCTCAGCTCAGCTCGTGTGTGTCTGCAATGCCACATA 180
QY 288 CTGTGACTCTTTGACCCCGACCTTCTTCCCTGCTGCTTCCAGCCCTATGAGAG 347
DB 181 CTGTGACTCTTTGACCCCGACCTTCTTCCCTGCTGCTTCCAGCCCTATGAGAG 240
QY 348 TACAGCAGTGGCGACGGATGGAGTCTGATGATGGGCCCCATCCAGGCTAATCACAGGG 407
DB 241 TACAGCAGTGGCGACGGATGGAGTCTGATGATGGGCCCCATCCAGGCTAATCACAGGG 300
QY 408 CACAGGCTGCTACTGACCTCCAGCCAGACAGAGAGTTCCAGAAAGTGAAGGATTTGG 467
DB 301 CACAGGCTGCTACTGACCTCCAGCCAGACAGAGAGTTCCAGAAAGTGAAGGATTTGG 360
QY 468 AGGGCCCATGACAGATGCTGTCTCTCAACATCTTCCCTGTCACCCCTGCCCAAAA 527
DB 361 AGGGCCCATGACAGATGCTGTCTCTCAACATCTTCCCTGTCACCCCTGCCCAAAA 420
QY 528 TTTGCTACTTAATCGTACTTCTCTGAAAGAGGAATCGGATATACATCATCCGGGTACC 587
DB 421 TTTGCTACTTAATCGTACTTCTCTGAAAGAGGAATCGGATATACATCATCCGGGTACC 480
QY 588 CATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGACAGACCCCTGATGATTT 647
DB 481 CATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGACAGACCCCTGATGATTT 540
QY 648 CCAGTTGGCACAACATTCAGCCTCCAGAGAGAGATACCAAGCTCAAGATACCCCTGATTC 707
DB 541 CCAGTTGGCACAACATTCAGCCTCCAGAGAGAGATACCAAGCTCAAGATACCCCTGATTC 600
QY 708 CCGAGCCCTGCAAGTTGGCCAGCTCCGTTTCTACTCTTCCAGCCCTCGGACATCAC 767
DB 601 CCGAGCCCTGCAAGTTGGCCAGCTCCGTTTCTACTCTTCCAGCCCTCGGACATCAC 660
QY 768 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGACAGCCCGG 827
DB 661 CACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTCAAGGACAGCCCGG 720
QY 828 AGACATCTACCAACAGCTGGGCGAGATACTTTGTGAAGTTCCTGGATGCTTGTCTGA 887

DB 721 AGACATCTACCACGACACCTGGGCCAGATACCTTTGTGAAGTTCTCGATGCCTATCTGA 780
QY 888 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAATAGACCTTCTGCTGGGCTGTTGAG 947
DB 781 GCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAATAGACCTTCTGCTGGGCTGTTGAG 840
QY 948 TGGATACCCCTTCCAGTGCCTGGGCTTACCCTGAAACATCAGCAGAGATTTCATTTGCCG 1007
DB 841 TGGATACCCCTTCCAGTGCCTGGGCTTACCCTGAAACATCAGCAGAGATTTCATTTGCCG 900
QY 1008 TGACCTAGGTCTTACCCTCGCCAAACAGTACTCACCAATGTCCGCTACTCATGTCTGA 1067
DB 901 TGACCTAGGTCTTACCCTCGCCAAACAGTACTCACCAATGTCCGCTACTCATGTCTGA 960
QY 1068 TGACCAACGCTTGTCTGCTCCCACTGGGCAAGGTGTACTGACAGACCCAGAAAGCAGC 1127
DB 961 TGACCAACGCTTGTCTGCTGCCCACTGGGCAAGGTGTACTGACAGACCCAGAAAGCAGC 1020
QY 1128 TAAATATGTTTCATGGGCTTGTCTGTACATTTGGTACCTGGACTTCTTGCTCCAGCCAAAGC 1187
DB 1021 TAAATATGTTTCATGGGCTTGTCTGTACATTTGGTACCTGGACTTCTTGCTCCAGCCAAAGC 1080
QY 1188 CACCTAGGGGAGACACACCGCTGTTCGCCAACACCATGCTCTTTGGCTCAGAGGCTTG 1247
DB 1081 CACCTAGGGGAGACACACCGCTGTTCGCCAACACCATGCTCTTTGGCTCAGAGGCTTG 1140
QY 1248 TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCTCTGGATCGAGGGATGA 1307
DB 1141 TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCTCTGGATCGAGGGATGA 1200
QY 1308 GTACAGCCACAGCATCATCAGAACCTCTGTACCATGTTGGTGGCTGACCGACTGGAA 1367
DB 1201 GTACAGCCACAGCATCATCAGAACCTCTGTACCATGTTGGTGGCTGACCGACTGGAA 1260
QY 1368 CCTTGGCTTGAACCCGGAAGGAGGACCCAAATTTGGTGGCTGCTTGTGACAGATCCCAT 1427
DB 1261 CCTTGGCTTGAACCCGGAAGGAGGACCCAAATTTGGTGGCTGCTTGTGACAGATCCCAT 1320
QY 1428 CATTTGATAGATCATCCAGAGACAGCTTTTAAACAGACCCATGTTTACACCTTGGCCA 1487
DB 1321 CATTTGATAGATCATCCAGAGACAGCTTTTAAACAGACCCATGTTTACACCTTGGCCA 1380
QY 1488 CTTTCCAGCAAGTTCTTCTCCTGAGGGCTCCCAAGAGTGGGCTGCTTGCAGTCAGAGAA 1547
DB 1381 CTTTCCAGCAAGTTCTTCTCCTGAGGGCTCCCAAGAGTGGGCTGCTTGCAGTCAGAGAA 1440
QY 1548 CGACCTGGACGCAAGTGGCACTGATGCCATCCCGATGGCTCTGTGTTGTGTGCTGTCTAAA 1607
DB 1441 CGACCTGGACGCAAGTGGCACTGATGCCATCCCGATGGCTCTGTGTTGTGTGCTGTCTAAA 1500
QY 1608 CGGCTCTCTAAGGATGTGGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTTGAGAC 1667
DB 1501 CGGCTCTCTAAGGATGTGGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTTGAGAC 1560
QY 1668 AATCTCACCTGGCTTACTTCCATTCCACACTACCTGTGGCTGCGCAGTGTGAGGACAGATA 1727
DB 1561 AATCTCACCTGGCTTACTTCCATTCCACACTACCTGTGGCTGCGCAGTGTGAGGACAGATA 1620
QY 1728 CTCAAGGAGGACTGGGCTCAGCTTGGGCTTAAAGGGACA 1768
DB 1621 CTCAAGGAGGACTGGGCTCAGCTTGGGCTTAAAGGGACA 1661

RESULT 2
AAA4844
ID AAA4844 standard; DNA; 1661 BP.
XX
XX AAA4844;
XX AC
XX DT 08-SEP-2000 (first entry)
XX DE Human lysosomal glycoprotein glucocerebrosidase (GCS) coding sequence.
XX

QY 858 CTTTGTGAAGTTCCTGGATGCTATGCTGAGACAAAGTTTACAGTTCTGGGAGTGACAGC 917
Db CTTTGTGAAGTTCCTGGATGCTATGCTGAGACAAAGTTTACAGTTCTGGGAGTGACAGC 848
QY 918 TGAATGAGCCCTTCTGCTGGGCTCTGAGTGATACCCCTCCAGTGCCTGGGCTTCAC 977
Db TGAATGAGCCCTTCTGCTGGGCTCTGAGTGATACCCCTCCAGTGCCTGGGCTTCAC 908
QY 978 CCTGTAACATCAGCAGACTTCATTGCTCCGCTGACCTAGCTAGCTCCTACCTCGCCACAGTAC 1037
Db CCTGTAACATCAGCAGACTTCATTGCTCCGCTGACCTAGCTAGCTCCTACCTCGCCACAGTAC 968
QY 1038 TCACACAAATGTCGCTTACTCATGCTGGATGACCAAGCTTGTGCTGCCCCACTGGGC 1097
Db TCACACAAATGTCGCTTACTCATGCTGGATGACCAAGCTTGTGCTGCCCCACTGGGC 1028
QY 1098 AAAGTGGTACTGACAGACCCAGAGCAGCTAAATATGTTATGCGATTGCTGTACATTG 1157
Db AAAGTGGTACTGACAGACCCAGAGCAGCTAAATATGTTATGCGATTGCTGTACATTG 1088
QY 1158 GTACCTGGACTTCTGGCTCCAGCCAAAGCCACCTAGGGAGACACACCGCTTGTCCC 1217
Db GTACCTGGACTTCTGGCTCCAGCCAAAGCCACCTAGGGAGACACACCGCTTGTCCC 1148
QY 1218 CAACACCATGCTTTGCTCAGAGCCCTGCTGGCTCCAAAGTTCTGGGAGCAGAGTGT 1277
Db CAACACCATGCTTTGCTCAGAGCCCTGCTGGCTCCAAAGTTCTGGGAGCAGAGTGT 1208
QY 1278 GGGCTGAGCTCTCTGGATGAGGAGTGCAGTACAGCCACACATCATCAGAACCTCT 1337
Db GGGCTGAGCTCTCTGGATGAGGAGTGCAGTACAGCCACACATCATCAGAACCTCT 1268
QY 1338 GTACCATGCTGCTGGCTGAGCCGACTGGAACCTTCCCTGAAACCCGAGGAGGCCAA 1397
Db GTACCATGCTGCTGGCTGAGCCGACTGGAACCTTCCCTGAAACCCGAGGAGGCCAA 1328
QY 1398 TTGGGTGCTTAATTTGTCGACAGTCCCATCATTTCTAGACATCACCAGGACAGTTTA 1457
Db TTGGGTGCTTAATTTGTCGACAGTCCCATCATTTCTAGACATCACCAGGACAGTTTA 1388
QY 1458 CAACAGCCCATGTTCTACACCTTGGCCATTTACAGCAAGTTTCACTTCTGAGGGCTCCCA 1517
Db CAACAGCCCATGTTCTACACCTTGGCCATTTACAGCAAGTTTCACTTCTGAGGGCTCCCA 1448
QY 1518 GAGAGTGGGCTGTTGTCAGTCAAGAACGACCTGGACGAGTGGCAGTATGCATCC 1577
Db GAGAGTGGGCTGTTGTCAGTCAAGAACGACCTGGACGAGTGGCAGTATGCATCC 1508
QY 1578 CGATGGCTCTGCTGTTGCTGCTGCTAAACCGCTCTTAAGGATGTCCTTTACCAT 1637
Db CGATGGCTCTGCTGTTGCTGCTGCTAAACCGCTCTTAAGGATGTCCTTTACCAT 1568
QY 1638 CAAGGATCTGCTGTTGGGCTTCTTGAGACAAATCTCACCTGGCTACTCCATTACACCTA 1697
Db CAAGGATCTGCTGTTGGGCTTCTTGAGACAAATCTCACCTGGCTACTCCATTACACCTA 1628
QY 1698 CTTGTCGCTGCTGATGATGAGCAGATCTCAAGAGGAGCTGGGCTCAGCCTGGGCA 1757
Db CTTGTCGCTGCTGATGATGAGCAGATCTCAAGAGGAGCTGGGCTCAGCCTGGGCA 1688
QY 1758 TTAAGGGACAGAGTCCAGC 1776
Db TTAAGGGACAGAGTCCAGC 1707

RESULT 5

ID AAN80113

XX AAN80113 standard; cDNA; 1661 BP.

AC AAN80113;

XX 17-DEC-2001 (updated)

DT 12-OCT-1990 (first entry)

XX Encodes recombinant human lysosomal glucocerebrosidase.
DE Lysosomal glycoprotein glucocerebrosidase ; Gaucher's disease; ss.
XX synthetic.
OS
XX Key Location/Qualifiers
FH 59..1607
CDS /*tag= a
FT /product=recombinant human lysosomal glucocerebrosidase
XX USN7137796-N.
XX 02-AUG-1988.
XX 23-DEC-1987; 87US-0192026.
XX 23-DEC-1987; 87US-0137796.
XX (USSH) US DEPT HEALTH AND HUMAN SERVICES.
XX Ginns EI;
XX WPI; 1988-264274/37.
DR P-PSDB; AAP80109.
XX Recombinant glucocerebrosidase -
PT reful for treating Gaucher's disease
XX Disclosure; ; 2pp; English.
XX The sequence is used for construction of the baculovirus derived
CC vector pAC373/GC. It differs from cDNA sequences published
CC previously, for example Leu (489) was previously given as Pro and
CC Arg (514) was previously given as His.
CC See also AAN80112.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NRI applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/atlis_us.html.)
XX SQ Sequence 1661 BP; 374 A; 496 C; 422 G; 369 T; 0 other;

Query Match 72.4%; Score 1649.8; DB 9; Length 1661;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1654; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 108 GGAGTCTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGTTAAGCATCAT 167
Db 1 GGAGTCTTCAAGTCTTCCAGAGAGGAATGTCCCAAGCCTTTGAGTAGGTTAAGCATCAT 60
QY 168 GCGTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCATCAGGTGCCCG 227
Db 61 GCGTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTCTGGGCATCAGGTGCCCG 120
QY 228 CCGCTGATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTCTGCAATGCCACATA 287
Db 121 CCGCTGATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTCTGCAATGCCACATA 180
QY 288 CTGTGACTCTCTTTCACCCCGACCTTTCCTTGGTGGTACCTTCAGCCGCTATGAGAG 347
Db 181 CTGTGACTCTCTTTCACCCCGACCTTTCCTTGGTGGTACCTTCAGCCGCTATGAGAG 240
QY 348 TACACGAGTGGGCGAGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGG 407
Db 241 TACACGAGTGGGCGAGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGG 300
QY 408 CACAGGCTGCTACTGACCCCTGCAGCCAGACAGAACTTCCAGAACTGAAGGATTGG 467
Db 301 CACAGGCTGCTACTGACCCCTGCAGCCAGACAGAACTTCCAGAACTGAAGGATTGG 360
QY 468 AGGGGCCATGACAGATGCTGCTCTCAACATCCTTTGCCCTGTACCCCTGCCAAAA 527

Db	1441	CGACCTGGACOCAGCTGGCCATTGATGATCATCCCGCATGGCTCTGCTGTGTGGTCGTCTAAA	1500
QY	1608	CGGCTCCCTCTAAGGATGTGCTCTTACCATCAAGATACCTGCTGTGGGCTTCTCTGGAGAC	1667
Db	1501	CGGCTCCCTCTAAGGATGTGCTCTTACCATCAAGATACCTGCTGTGGGCTTCTCTGGAGAC	1560
QY	1668	AATCTCACCTGGCTACTTCCATTCACACTACTCTGTGGCTCGGCACGTGATGGAGCAGATA	1727
Db	1561	AATCTCACCTGGCTACTTCCATTCACACTACTCTGTGGCTCGGCACGTGATGGAGCAGATA	1620
QY	1728	CTCAAGAGGCACCTGGGCTCAGCCTGGGCATTAAAGGGACA	1768
Db	1621	CTCAAGAGGCACCTGGGCTCAGCCTGGGCATTAAAGGGACA	1661

RESULT 6	
AA71753	
ID	AA71753 standard; cDNA; 1642 BP.
XX	
XX	AA71753;
XX	
XX	01-OCT-1997 (first entry)
XX	
XX	Human glucocerebrosidase-FLAG cDNA construct.
DE	
XX	
XX	Lysosomal enzyme; lysosome; transgenic plant; glucocerebrosidase;
KW	enzyme replacement therapy; Gaucher disease; ss.
KW	
XX	
XX	Chimaeric Homo sapiens;
OS	Chimaeric synthetic.
OS	
XX	
XX	Key
XX	Location/Qualifiers
FT	1..1642
FT	/*tag= a
FT	/transl_except= (pos:265..267, aa:Glu)
XX	
XX	W09710353-A1.
XX	
XX	
XX	20-MAR-1997.
PD	
XX	
XX	13-SEP-1996; 96WO-US14730.
XX	
XX	
XX	14-SEP-1995; 95US-0003737.
XX	
XX	(CROP-) CROPTech DEV CORP.
PA	(VIRG) VIRGINIA TECH INTELLECTUAL PTY INC.
PA	
PI	
PI	Cramer CL, Oishi KK, Radin DN, Weissenborn DL;
DR	
DR	WPI; 1997-202248/18.
DR	P-PSDB; AA718237.
XX	
PT	Production of enzymatically active (modified) lysosomal enzyme in
PT	transgenic plants - useful in treatment of lysosomal storage
PT	disorders
XX	
XX	Example 1; Page 61-62; 111pp; English.
XX	
XX	A cDNA construct (AA71753) codes for a fusion protein (AA718237) comprising human glucocerebrosidase (hGC) and a C-terminal FLAG epitope (see also AA718236). The hGC sequence was obt'd. by PCR amplification using primers (see also AA71755-56) designed to incorporate restriction sites to facilitate cloning and ligation of the FLAG coding sequence. The construct, placed under control of the inducible MeaA promoter (see also AA71752), can drive expression of hGC in transgenic (esp. tobacco) plants. The plant expression system provides for post-translational modification and processing to produce enzymatically active hGC, which can be harvested for use in Gaucher disease enzyme replacement therapy.

Query Match	Score 1601.6;	DB 18;	Length 1642;
70.3%			

Best local similarity	99.8%	pred	No. 0:
-----------------------	-------	------	--------

BEST LOCAL SIMILARITY 99.8%, P-ED: NO: 0,
Matches 1604: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy	106	ATGGAGTTTTC	AAAGTCCCTT	CCAGAGAGGAAT	TGTC	CCAAAGCCCTT	TGAGTACGGTAA	GCATC	155									
Db	1	ATGGAGTTTTC	AAAGTCCCTT	CCAGAGAGGAAT	TGTC	CCAAAGCCCTT	TGAGTACGGTAA	GCATC	60									
Qy	166	ATGGCTGGCAG	CGCTCACAGGAT	TGCTTCTACTT	CAGGCAGT	GTGCTGGGCAT	CAGGTGCC	225										
Db	61	ATGGCTGGCAG	CGCTCACAGGAT	TGCTTCTACTT	CAGGCAGT	GTGCTGGGCAT	CAGGTGCC	120										
Qy	226	CGCCCCGTGCAT	CCCTAAAGCTT	CGGCTACAGCT	CGGTGGTGGTGTCT	GCAATGCAC	CA	285										
Db	121	CGCCCCGTGCAT	CCCTAAAGCTT	CGGCTACAGCT	CGGTGGTGGTGTCT	GCAATGCAC	CA	180										
Qy	286	TACTGTGACT	CTCTTTGACCC	CCCGACCTT	CTCTGCCCC	TGCTTCA	CAGCCGTATGAG	345										
Db	181	TACTGTGACT	CTCTTTGACCC	CCCGACCTT	CTCTGCCCC	TGCTTCA	CAGCCGTATGAG	240										
Qy	346	AGTACACG	CAGTGGCGAG	CGGATGAGT	ATG	GGGCCAT	TCCAGGCTTA	TCACACG	405									
Db	241	AGTACACG	CAGTGGCGAG	CGGATGAGT	ATG	GGGCCAT	TCCAGGCTTA	TCACACG	300									
Qy	406	GGCACAG	CGCTCTACT	GACCTCG	CAGCAG	ACAAGAGT	TCCAGA	AGTGAAGGATTT	465									
Db	301	GGCACAG	CGCTCTACT	GACCTCG	CAGCAG	ACAAGAGT	TCCAGA	AGTGAAGGATTT	360									
Qy	466	GGAGGGCCAT	GACAGAT	CTGCTCTCT	CAACAT	TCCTTGCCCT	TGTCA	CCCCCTGCCCAA	525									
Db	361	GGAGGGCCAT	GACAGAT	CTGCTCTCT	CAACAT	TCCTTGCCCT	TGTCA	CCCCCTGCCCAA	420									
Qy	526	AATTTGCT	ACTTAAT	CGTACTTCT	CTG	AGAGGAAG	AATCGGAT	ATAACATCAT	CCCGGTA	585								
Db	421	AATTTGCT	ACTTAAT	CGTACTTCT	CTG	AGAGGAAG	AATCGGAT	ATAACATCAT	CCCGGTA	480								
Qy	586	CCCATGCC	AGCTGTGACT	TTCCAT	TCGGC	ACTAC	ACTATG	CAGACAC	CCCCCTGATGAT	645								
Db	481	CCCATGCC	AGCTGTGACT	TTCCAT	TCGGC	ACTAC	ACTATG	CAGACAC	CCCCCTGATGAT	540								
Qy	646	TTCCAGTT	TGCAC	NACTT	CAGCCT	CCAGAG	GAGAT	ACCAAGCT	CAAGATAC	CCCTGATT	705							
Db	541	TTCCAGTT	TGCAC	NACTT	CAGCCT	CCAGAG	GAGAT	ACCAAGCT	CAAGATAC	CCCTGATT	600							
Qy	706	CACCGAG	CCCTGCAG	TGGCC	CAGCGT	CCCGTTT	CAC	TCCTTGGC	CAGCCCTG	GACATCA	765							
Db	601	CACCGAG	CCCTGCAG	TGGCC	CAGCGT	CCCGTTT	CAC	TCCTTGGC	CAGCCCTG	GACATCA	660							
Qy	766	CCCATTTGG	CTAAG	ACAAT	TGAG	CGGTGA	ATGGGA	AGGGT	CAC	TCAAGGGAC	825							
Db	661	CCCATTTGG	CTAAG	ACAAT	TGAG	CGGTGA	ATGGGA	AGGGT	CAC	TCAAGGGAC	720							
Qy	826	GGAGACAT	CTTACC	CAC	CAGACCT	TGGG	CCAGAT	ACTTTGT	TGAAGTT	TCCTGGAT	885							
Db	721	GGAGACAT	CTTACC	CAC	CAGACCT	TGGG	CCAGAT	ACTTTGT	TGAAGTT	TCCTGGAT	780							
Qy	886	GAGCACA	AGTTTAC	AGTTCT	TGGG	CAGTG	CAGCTG	AAAAAT	TAG	CCCTTCTG	CTGGGCTTTG	945						
Db	781	GAGCACA	AGTTTAC	AGTTCT	TGGG	CAGTG	CAGCTG	AAAAAT	TAG	CCCTTCTG	CTGGGCTTTG	840						
Qy	946	AGTGGAT	TACCCCTT	CCAGT	GCCTT	CGGCTT	CA	CCCTTGA	ACAT	CAGC	GAGCTT	CA	1005					
Db	841	AGTGGAT	TACCCCTT	CCAGT	GCCTT	CGGCTT	CA	CCCTTGA	ACAT	CAGC	GAGCTT	CA	900					
Qy	1006	CGTGAC	TAGGTCCT	TACCCT	TCGCC	ACAAGT	ACT	CA	CCACAAT	TGCC	CCCTACT	CA	1065					
Db	901	CGTGAC	TAGGTCCT	TACCCT	TCGCC	ACAAGT	ACT	CA	CCACAAT	TGCC	CCCTACT	CA	960					
Qy	1066	GATGACCA	ACGCTT	GTCTGCT	CCCC	ACTGGG	CAAG	GTG	ACT	CA	CAGAC	CCCC	AGA	1125				
Db	961	GATGACCA	ACGCTT	GTCTGCT	CCCC	ACTGGG	CAAG	GTG	ACT	CA	CAGAC	CCCC	AGA	1020				
Qy	1126	GCTAAAT	TATTT	CA	TGGCAT	TGCTGT	TACAT	TGGT	TAC	CT	ACCT	TG	AGCTT	TTCTGG	CT	CCAG	CCAAA	1185

Db	1021	GCTAAATATCTTCATGGCAATTGCTGTACATTGGTACCTTGGACATTTCTGGCTTCAGACCAA	1080
Qy	1186	GCACACCTAGGGAGACACACCGCCTGTTTCCCAACACCATGCTCTTTGGCTCAGAGGCC	1245
Db	1081	GCACACCTAGGGAGACACACCGCCTGTTTCCCAACACCATGCTCTTTGGCTCAGAGGCC	1140
Qy	1246	TGTTGGGCTCCAAAGTTCTTGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATG	1305
Db	1141	TGTTGGGCTCCAAAGTTCTTGGAGCAGAGTGTGGGCTAGGCTCCTGGGATCGAGGGATG	1200
Qy	1306	CAGTACAGCCACAGATCATCAGCAACTCCTGTGTACCATGTGGTGGCTGAGACGACTGG	1365
Db	1201	CAGTACAGCCACAGATCATCAGCAACTCCTGTGTACCATGTGGTGGCTGAGACGACTGG	1260
Qy	1366	AACCTTGCCTGAAACCCGGAAGGAGGACCAATTGGTGCGTAACTTTGTGCGACAGTCCC	1425
Db	1261	AACCTTGCCTGAAACCCGGAAGGAGGACCAATTGGTGCGTAACTTTGTGCGACAGTCCC	1320
Qy	1426	ATCATTTAGACATCATCCAAAGGACAGCTTTTACAACAGSCCATGTTCTTACCACCTTTGGC	1485
Db	1321	ATCATTTAGACATCATCCAAAGGACAGCTTTTACAACAGSCCATGTTCTTACCACCTTTGGC	1380
Qy	1486	CAC TTCAGAAAGTTCAATTCCTGAGGCTCCCAAGAGATGGGGCTGGTGCACAGTCAGAAG	1545
Db	1381	CAC TTCAGAAAGTTCAATTCCTGAGGCTCCCAAGAGATGGGGCTGGTGCACAGTCAGAAG	1440
Qy	1546	AACGACCTGCACGACGTGGCAGTCATGCCATCCGATGGCTCTGCTGTGGTGTGCTGCTGCTA	1605
Db	1441	AACGACCTGCACGACGTGGCAGTCATGCCATCCGATGGCTCTGCTGTGGTGTGCTGCTGCTA	1500
Qy	1606	AACGGCTCCTCTAAGGATGTGCCTCTTACCATCAAGGATTCCTGCTGTGGGCTTCCTGGAG	1665
Db	1501	AACGGCTCCTCTAAGGATGTGCCTCTTACCATCAAGGATTCCTGCTGTGGGCTTCCTGGAG	1560
Qy	1666	ACAAATCTCACCTGGCTACTCCATTCACACTACCTGTGGCGTGCACG	1713
Db	1561	ACAATCTCACCTGGCTACTCCATTCACACTACCTGTGGCGTGCACG	1608

RESULT 7

RESOLUT
AAD06389

AA006389
ID AA006389 standard: cDNA: 1611 BP.

AA AAD06389:

DT 10-AUG-2001 (first entry)

Human glucocerebrosidase (GC) cDNA #2.

Human; adeno-associated viral expression vector; AAV; gene therapy;
KW
KW lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
KW Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG;
KW synoviarthroblastosis; GC; Gaucher's disease; ss.

XX.
OS.
Homo sapiens.

AA	Key	Location/Qualifiers
FH		1.1611
FT	CDS	

FT /product=

```

F'P
b=3-faces
/*tag= b

```

```
FT
mscfttag= c
/*tag= c
```

XX
1,

XX XX

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

XX

1000

30-JUN-2000; 2000US-0215430.
(AVIG-) AVIGEN INC.
(CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
Podsakoff G, Watson G, Couto LB, Yang B;
WPI; 2001-343814/36.
P-PSDB; AAE02446.
Use of recombinant adeno-associated virus, comprising gene encoding a protein defective or missing in lysosomal storage disease, in the manufacture of a medicament for treating the lysosomal storage disease
Example 3a; Page 91-94; 97pp; English.
The present invention relates to recombinant adeno-associated virus (AAV) expression vectors and virions, which include genes coding for enzymes defective or missing in lysosomal storage disease (LSD). AAV is useful in the manufacture of a medicament for treating lysosomal storage disease e.g., mucopolysaccharidoses VII (MPS VII), MPS VII (Sly Syndrome) which aids in the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) gene therapy. The present sequence is human glucocerebrosidase (GC) cDNA. Mutations in GC gene leads to Gaucher's disease, a lysosomal storage disease. This sequence is used in AAV constructs.
Sequence 1611 BP; 359 A; 487 C; 404 G; 361 T; 0 other;
Query Match 69.6%; Score 1587; DB 22; Length 1611;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1596; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
106 ATGGAGTTTCAAGTCTTCCAGAGAGGAATGTCCTCCAGAGCTTTGAGTGAAGCATC 165
1 ATGGAGTTTCAAGTCTTCCAGAGAGGAATGTCCTCCAGAGCTTTGAGTGAAGCATC 60
166 ATGGCTGGCAGCCTCACAGGATGTTCTACTTCAGGCAGTGCCTGGGCATCAGGTGCC 225
61 ATGGCTGGCAGCCTCACAGGATGTTCTACTTCAGGCAGTGCCTGGGCATCAGGTGCC 120
226 CGCCCTTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 285
121 CGCCCTTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACA 180
286 TACTGTGACTCCTTTGACCCCGGACCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
181 TACTGTGACTCCTTTGACCCCGGACCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
346 AGTACAGCAGTGGCGAGGATGAGCTGATGAGTGGGCGGATGAGTGGGCGGATGAGTGG 405
241 AGTACAGCAGTGGCGAGGATGAGCTGATGAGTGGGCGGATGAGTGGGCGGATGAGTGG 300
406 GGCACAGCCTGCTACTGACCTGACGCGCAGCAGAGGATTCAGAAAGTGAAGGATTT 465
301 GGCACAGCCTGCTACTGACCTGACGCGCAGCAGAGGATTCAGAAAGTGAAGGATTT 360
466 GGAGGGCCATGACAGATGCTGCTCTCAACATCCTTCCCTGTCAACCTGCTGCTGCTGCT 525
361 GGAGGGCCATGACAGATGCTGCTCTCAACATCCTTCCCTGTCAACCTGCTGCTGCTGCT 420
526 AATTGCTACTTAAATCGTACTTCTCTGAAGAGGAATTCGATATTAACATATCCTCGGATA 585
421 AATTGCTACTTAAATCGTACTTCTCTGAAGAGGAATTCGATATTAACATATCCTCGGATA 480
586 CCCATGCCAGTGTGACTTCTCTCAACCTGACCTATGACAGACACCTGATGAT 645
481 CCCATGCCAGTGTGACTTCTCTCAACCTGACCTATGACAGACACCTGATGAT 540
646 TTCAGTTGCACAACTTACAGCTTCCAGAGGATACCAAGCTCAAGATACCCCTGATT 705
541 TTCAGTTGCACAACTTACAGCTTCCAGAGGATACCAAGCTCAAGATACCCCTGATT 600

QY 706 CACGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTTGCAGCCCTTGACATCA 765
DB 601 CACGAGCCCTGCAGTTGGCCAGCGTCCCGTTTCACTCTTGCAGCCCTTGACATCA 660
QY 766 CCCACTTGGCTCAAGACCAATGAGGCGGTGAATGGAAGGGGTCACTCAAGGACAGCCC 825
DB 661 CCCACTTGGCTCAAGACCAATGAGGCGGTGAATGGAAGGGGTCACTCAAGGACAGCCC 720
QY 826 GGAGACATCTACCAACAGACTGGGCGAGATCTTTGTGAAGTTCTCTGGATGCTATGCT 885
DB 721 GGAGACATCTACCAACAGACTGGGCGAGATCTTTGTGAAGTTCTCTGGATGCTATGCT 780
QY 886 GAGCAACAAGTTTACAGTTCTGGGCGAGTGACAGCTGAAATGAGCCCTTCTGGGCTGTTG 945
DB 781 GAGCAACAAGTTTACAGTTCTGGGCGAGTGACAGCTGAAATGAGCCCTTCTGGGCTGTTG 840
QY 946 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCTGACATGACGAGACTTCATGCT 1005
DB 841 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCTGACATGACGAGACTTCATGCT 900
QY 1006 CGTACCTAGTCTCTACCTCGCCACACAGTACTACACCAATGTCCGCTACTCATGCTG 1065
DB 901 CGTACCTAGTCTCTACCTCGCCACACAGTACTACACCAATGTCCGCTACTCATGCTG 960
QY 1066 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
DB 961 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1126 GCTAAATATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1185
DB 1021 GCTAAATATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
QY 1186 GCACCCCTAGGGGAGACACACCGCTGTTCCCAACACCATGCTCTTTGCTGCTGCTGCTG 1245
DB 1081 GCACCCCTAGGGGAGACACACCGCTGTTCCCAACACCATGCTCTTTGCTGCTGCTGCTG 1140
QY 1246 TGTGTGGGCTCCAGTTCTGGGAGAGTGTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1305
DB 1141 TGTGTGGGCTCCAGTTCTGGGAGAGAGTGTGGGCTGAGGCTGAGGCTGAGGCTGAGG 1200
QY 1306 CAGTACAGCCACGATCATCAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365
DB 1201 CAGTACAGCCACGATCATCAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1366 AACCTTGGCTGAGACCCGAGAGGACCAATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1425
DB 1261 AACCTTGGCTGAGACCCGAGAGGACCAATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1426 ATCATTTGAGACATCAGACAGGACAGTTTACAAAGCCATGTTTCTACCACTTGGC 1485
DB 1321 ATCATTTGAGACATCAGACAGGACAGTTTACAAAGCCATGTTTCTACCACTTGGC 1380
QY 1486 CACTTCAGCAAGTTTCAATCTTCTGAGGCTCCAGAGAGTGGGCTGCTGCTGCTGCTGCTGCTG 1545
DB 1381 CACTTCAGCAAGTTTCAATCTTCTGAGGCTCCAGAGAGTGGGCTGCTGCTGCTGCTGCTGCTG 1440
QY 1546 AACGACCTGGACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1605
DB 1441 AACGACCTGGACGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 1606 AACGCTCTCTTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
DB 1501 AACGCTCTCTTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
QY 1666 ACAATCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1716
DB 1561 ACAATCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1611
RESULT 8
AAD06388
ID AAD06388 standard; cDNA; 1551 BP.

XX AAD06388;
XX 10-AUG-2001 (first entry)
XX Human glucocerebrosidase (GC) cDNA #1.
XX Human; adeno-associated viral expression vector; AAV; gene therapy;
KW lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
KW Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG;
KW glucocerebrosidase; GC; Gaucher's disease; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1..1551
FT /*tag= a
FT /product= "Human glucocerebrosidase (GC)"
FT sig_peptide 1..57
FT /*tag= b
FT mat_peptide 58..1548
FT /*tag= c
FT /product= "Human mature glucocerebrosidase (GC)"
XX WO200136603-A2.
XX 25-MAY-2001.
XX 17-NOV-2000; 2000WO-US31688.
XX 17-NOV-1999; 99US-0166097.
PR 30-JUN-2000; 2000US-0215430.
XX (AVIG-) AVIGEN INC.
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX Podsakoff G, Watson G, Couto LB, Yang B;
PI WPI; 2001-343814/36.
XX P-PSDB; AAE02445.
XX Use of recombinant adeno-associated virus, comprising a
PT protein defective or missing in lysosomal storage disease, in the
PT manufacture of a medicament for treating the lysosomal storage disease
PT .
XX Example 3A; Page 87-89; 97pp; English.
XX The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful
CC in the manufacture of a medicament for treating lysosomal storage
CC disease e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome)
CC is due to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS)
CC which aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in
CC gene therapy. The present sequence is human glucocerebrosidase (GC)
CC cDNA. Mutations in GC gene leads to Gaucher's disease, a lysosomal
CC storage disease. This sequence is used in AAV constructs.
XX Sequence 1551 BP; 344 A; 474 C; 387 G; 346 T; 0 other;
Query Match 67.9%; Score 1547.8; DB 22; Length 1551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 166 ATGGCTGGCAGCTCACAGATTGCTTCTACTTCAGGAGTGTCTGGGCGATCAGGTGCC 225
DB 1 ATGGCTGGCAGCTCACAGATTGCTTCTACTTCAGGAGTGTCTGGGCGATCAGGTGCC 60
QY 226 CGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTCTGCAATGCCACA 285
DB 61 CGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTCTGCAATGCCACA 120

QY 286 TACTGTGACTCTCTTTGACCCCCCGACCTTTCTGSCCCTTGGTACCTTACGCCCTATGAG 345
DB 121 TACTGTGACTCTCTTTGACCCCCCGACCTTTCTGSCCCTTGGTACCTTACGCCCTATGAG 180
QY 346 AGTACACGAGTGGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATACACG 405
DB 181 AGTACACGAGTGGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATACACG 240
QY 406 GGCACAGCCTGTCTACTGACCTTCAGCCAGACAGAAAGTTCCAGAAAGTGAAGGGATT 465
DB 241 GGCACAGCCTGTCTACTGACCTTCAGCCAGACAGAAAGTTCCAGAAAGTGAAGGGATT 300
QY 466 GGAGGGCCATGACAGATGTCTGCTCTCAACATCTTGGCCTGTCCACCCCTGCCCAA 525
DB 301 GGAGGGCCATGACAGATGTCTGCTCTCAACATCTTGGCCTGTCCACCCCTGCCCAA 360
QY 526 AATTGTGTACTTAATCGTACTTCTCTGAAGAAGAAATCGATATAACATCATCCGGTA 585
DB 361 AATTGTGTACTTAATCGTACTTCTCTGAAGAAGAAATCGATATAACATCATCCGGTA 420
QY 586 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACACCCCTGATGAT 645
DB 421 CCCATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGGCACAACTTCAGCCTCCAGAGAGAAATACCAAGCTCAAGATACCCCTGATT 705
DB 481 TTCCAGTTGGCACAACTTCAGCCTCCAGAGAGAAATACCAAGCTCAAGATACCCCTGATT 540
QY 706 CACGAGCCCTGCAGTTGGCCCGACGCTCCCTTTCACTTCTTGGCAGCCCTGGACATCA 765
DB 541 CACGAGCCCTGCAGTTGGCCCGACGCTCCCTTTCACTTCTTGGCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTAAGGGACAGCCC 825
DB 601 CCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGGTCACTAAGGGACAGCCC 660
QY 826 GGAGACATCTACCACAGACCTGGGCCAGATCTTTGTGAAGTTCTTGGATGCCATGCT 885
DB 661 GGAGACATCTACCACAGACCTGGGCCAGATCTTTGTGAAGTTCTTGGATGCCATGCT 720
QY 886 GAGCAAGTTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGTGGCTGTG 945
DB 721 GAGCAAGTTTACAGTTCTGGGCAGTGACAGCTGAAATGAGCCTTCTGTGGCTGTG 780
QY 946 AGTGATACCCCTTCCAGTGCTGGGCTTACCCCTGACATCAGCGAGACTTCATTGCC 1005
DB 781 AGTGATACCCCTTCCAGTGCTGGGCTTACCCCTGACATCAGCGAGACTTCATTGCC 840
QY 1006 CGTGACCTAGTCTTACCTCCCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 1065
DB 841 CGTGACCTAGTCTTACCTCCCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 900
QY 1066 CATGACCAACGCTTGTCTGCTGCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCA 1125
DB 901 GATGACCAACGCTTGTCTGCTGCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCA 960
QY 1126 GCTAAATATGTTATGGATGTCTGATGTTGTTACCTGGACTTCTGGCTCCAGCCAAA 1185
DB 961 GCTAAATATGTTATGGATGTCTGATGTTGTTACCTGGACTTCTGGCTCCAGCCAAA 1020
QY 1186 GCACCCCTAGGGGAGACACACCGCTGTTCGCCAACACACATGCTTTCCTCCAGAGGCC 1245
DB 1021 GCACCCCTAGGGGAGACACACCGCTGTTCGCCAACACACATGCTTTCCTCCAGAGGCC 1080
QY 1246 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGGCTAGGCTCTCTGGGATCGAGGATG 1305
DB 1081 TGTGTGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGGCTAGGCTCTCTGGGATCGAGGATG 1140
QY 1306 CAGTACACCCAGCATCATCAGCAACCTCTGTACCAATGTGGTGGCTGACCACTGG 1365
DB 1141 CAGTACACCCAGCATCATCAGCAACCTCTGTACCAATGTGGTGGCTGACCACTGG 1200
QY 1366 AACCTTGGCCCTGAACCCCGAAGGAGGCCCAATTTGGGTGCGTAACTTTTGTGACAGATCCC 1425

Db 1201 AACCTTGCCCTGACCCGAGGAGGAGCCCAATTTGGTGGTAACTTTGTCACAGTCCC 1260
 Qy 1426 ATCATTTGTAGACATACCAAGACACGTTTACAAAGCCCAATGTTTACACCTTGGC 1485
 Db 1261 ATCATTTGTAGACATACCAAGACACGTTTACAAAGCCCAATGTTTACACCTTGGC 1320
 Qy 1486 CACTTCAGCAAGTTATCTTCTGAGGGTCCAGAGAGTGGGGTGGTGGCCAGTCAAG 1545
 Db 1321 CACTTCAGCAAGTTATCTTCTGAGGGTCCAGAGAGTGGGGTGGTGGCCAGTCAAG 1380
 Qy 1546 AAGGACCTGGAGCCAGTGGCACTGATGCATCCCGATGGCTCTGCTTTGTTGGTCTGCTA 1605
 Db 1381 AAGGACCTGGAGCCAGTGGCACTGATGCATCCCGATGGCTCTGCTTTGTTGGTCTGCTA 1440
 Qy 1606 AAGCGTCTCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAG 1665
 Db 1441 AAGCGTCTCTTAAGGATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTGGAG 1500
 Qy 1666 ACAATCTCACTGGCTACTCCATTCACACCTACCTGTGGGTCGCCAGTGA 1716
 Db 1501 ACAATCTCACTGGCTACTCCATTCACACCTACCTGTGGGTCGCCAGTGA 1551

RESULT 9
 AAH99925 standard; cDNA; 1593 BP.
 XX AC AAH99925;
 XX DT 13-FEB-2002 (first entry)
 XX DE Nucleotide sequence of human glucocerebrosidase.
 XX KW Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
 KW Farber disease; Tay-sachs disease; Niemann-pick disease;
 KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
 KW Gaucher disease; glycosylation; ss.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 XX CDS 13..1563
 XX FT /*tag= a
 XX FT /product= "Human glucocerebrosidase"
 XX PN W0200177307-A2.
 XX PD 18-OCT-2001.
 XX PF 06-APR-2001; 2001WO-US11144.
 XX PR 06-APR-2000; 2000US-195598P.
 XX PA (CYTO-) CYTOCLONAL PHARM INC.
 XX PI Berent SL;
 XX PI WPI; 2002-041292/05.
 XX DR P-PSDB; AAG78411.
 XX PT Novel expression system useful for producing glucocerebrosidase for
 PT treating Gaucher's disease, comprises an insect cell transformed with a
 PT vector encoding glucocerebrosidase that synthesises glucocerebrosidase
 PT .
 XX Claim 3; Page 67-69; 74pp; English.
 XX CC This invention relates to an expression system comprising an insect
 CC cell transformed with a vector encoding glucocerebrosidase (GC) that
 CC synthesises clinically effective GC. Recombinant GC is useful for
 CC treating individuals with deficiencies in GC, by introducing the
 CC clinically effective form of recombinant GC from insect cells into

CC individuals. The expression system is stably transformed and
 CC provides a consistently higher level of expression of GC than in
 CC baculovirus or mammalian cell expression systems, and proper
 CC glycosylation modifications for GC, requiring no enzymatic
 CC carbohydrate remodeling to be clinically effective. The expression
 CC system is more effective, economical and simple for the manufacture of
 CC of lysosomal enzymes contributes to some of the many genetically
 CC inherited lysosomal storage diseases, such as Fabry disease, Farber
 CC disease, Tay-Sachs disease, Niemann-pick disease, Hunter syndrome,
 CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease
 CC This nucleotide sequence represents that of human
 CC glucocerebrosidase. See also AAH99924 for sequence of human
 CC glucocerebrosidase, noting the differences at residues 1552 to 1554.
 CC This sequence encoded histidine at that position, AAH99924 encodes
 CC arginine.
 XX
 SQ Sequence 1593 BP; 359 A; 484 C; 396 G; 354 T; 0 other;
 Query Match 67.7%; Score 1544; DB 24; Length 1593;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1550; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Qy 157 GTAAGCATCATGGCTGGCAGCCTCACAGGATTTGTTTACTTCAGGCAGTGTGTTGGGCA 216
 Db 4 GAATTTCAAATGGCTGGCAGCCTCACAGGTTTGTCTTACTTCAGGCAGTGTGTTGGGCA 63
 Qy 217 TCAGGTGCCCGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGC 276
 Db 64 TCAGGTGCCCGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGC 123
 Qy 277 AATGCCACATACGTGACTCTTTCAGCCGCCGACCTTTCTCCCTTGTACTTACCTTCAGC 336
 Db 124 AATGCCACATACGTGACTCTTTCAGCCGCCGACCTTTCTCCCTTGTACTTACCTTCAGC 183
 Qy 337 CGCTATGAGATACACGAGTGGGGAGGATGGAGCTGAGTATGGGCCCATCCAGCT 396
 Db 184 CGCTATGAGATACACGAGTGGGGAGGATGGAGCTGAGTATGGGCCCATCCAGCT 243
 Qy 397 AATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGCAAGAGTTCCAGAAAGTG 456
 Db 244 AATCACACGGGCACAGGCTGCTACTGACCTGCAGCCAGCAAGAGTTCCAGAAAGTG 303
 Qy 457 AAGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCTTGCCTGTCAACC 516
 Db 304 AAGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCTTGCCTGTCAACC 363
 Qy 517 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 576
 Db 364 CCTGCCCAAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATC 423
 Qy 577 ATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACCC 636
 Db 424 ATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTTACACCTATGCAGACCC 483
 Qy 637 CCTGATGATTTCCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATA 696
 Db 484 CCTGATGATTTCCAGTTGCACAACTTCAGCCCTCCAGAGGAAGATACCAAGCTCAAGATA 543
 Qy 697 CCCTGATTTACCGAGCCCTGCGAGTTGGCCAGGTCCTGCTTCACTCTTGGCGAGCCC 756
 Db 544 CCCTGATTTACCGAGCCCTGCGAGTTGGCCAGGTCCTGCTTCACTCTTGGCGAGCCC 603
 Qy 757 TGCATATACCCCTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGGTCACTCAAG 816
 Db 604 TGCATATACCCCTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGGTCACTCAAG 663
 Qy 817 GGACAGCCCGGAGACATCTTACACACAGACCTTGGGCCAGATACCTTGTGAAGTTCCTGGAT 876
 Db 664 GGACAGCCCGGAGACATCTTACACACAGACCTTGGGCCAGATACCTTGTGAAGTTCCTGGAT 723
 Qy 877 GCCTATGCTGAGCACAAAGTTACAGTTCTGGGCACTGACAGCTGAAATGAGCTTCTGCT 936

QY	526	AA	TTTGCTACTTAAATCGTACTTCTCGAAGAGGAATCGGATATACATCATCCGGGTA	595
DB	361	AA	TTTTGCTACTTAAATCGTACTTCTCGAAGAGGAATCGGATATAAATCATCATCCGGGTA	420
QY	586	CC	ATGCGCAGCTGTGACTTCTCCATCGCGACCTTACACCTATGCGACACCCCGTGATGAT	645
DB	421	CC	ATGCGCAGCTGTGACTTCTCCATCGCGACCTTACACCTATGCGACACCCCGTGATGAT	480
QY	646	TT	CCAGTGTGCACAACATTCAGCTCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT	705
DB	481	TT	CCAGTGTGCACAACATTCAGCTCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT	540
QY	706	CAC	GAGCCCTCGAGTGTGGCCAGCGTCCCGGTTTTCACCTCTTGCAGGCCCTTGACATCA	765
DB	541	CAC	GAGCAGTGCAGTGTGGCCAGCGTCCCGGTTTTCACCTCTTGCAGGCCCTTGACATCA	600
QY	766	CC	ACTTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGSGACAGCCC	825
DB	601	CC	ACTTTGGCTCAAGACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGSGACAGCCC	660
QY	826	GG	GACATCTACCAACAGACCTGGCGAGATACTTGTGAAGTTCCTGGATGCCATGCT	885
DB	661	GG	GACATCTACCAACAGACCTGGCGAGATACTTGTGAAGTTCCTGGATGCCATGCT	720
QY	886	GAG	CACAAGTTACAGTTCGTGGCAGTGACAGCTGAATAGCCTTCTGCTGGGCTCTTG	945
DB	721	GAG	CACAAGTTACAGTTCGTGGCAGTGACAGCTGAATAGCCTTCTGCTGGGCTCTTG	780
QY	946	AG	TGGATACCCCTTCAGATGCCGTGGGGCTTACCCCTGAAATCAGCGAGAGCTTCAATGCC	1005
DB	781	AG	TGGATACCCCTTCAGATGCCGTGGGGCTTACCCCTGAAATCAGCGAGAGCTTCAATGCC	840
QY	1006	CG	TGACCTTAGTGCTTACCCTCGCCAAACAGTACTCACACAATGTCCGCTACTCATGCTG	1065
DB	841	CG	TGACCTTAGTGCTTACCCTCGCCAAACAGTACTCACACAATGTCCGCTACTCATGCTG	900
QY	1066	GAT	GACCAACGCTTGCTGCTGCCACACTGGGCAAGGTGTACTGACAGACCCAGAAGCA	1125
DB	901	GAT	GACCAACGCTTGCTGCTGCCACACTGGGCAAGGTGTACTGACAGACCCAGAAGCA	960
QY	1126	GCT	AAATATCTTATGGCATTCGTGATCATGTGTACTGGAGCTTTCTGGCTCGAGCCAAA	1185
DB	961	GCT	AAATATCTTATGGCATTCGTGATCATGTGTACTGGAGCTTTCTGGCTCGAGCCAAA	1020
QY	1186	GC	ACCTTAGGGGAGACACACCGCTGTTCGCCAACACATGCTTTTGCTTCAGAGCC	1245
DB	1021	GC	ACCTTAGGGGAGACACACCGCTGTTCGCCAACACATGCTTTTGCTTCAGAGCC	1080
QY	1246	TG	TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCTTGGGATCGAGGGATG	1305
DB	1081	TG	TGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCTTGGGATCGAGGGATG	1140
QY	1306	CAG	TACAGCCACAGCATCATCAGAAACCTCGTGTACCATGTGGTGGCTGGACCGACTGG	1365
DB	1141	CAG	TACAGCCACAGCATCATCAGAAACCTCGTGTACCATGTGGTGGCTGGACCGACTGG	1200
QY	1366	AA	CTTGTGCTGAACCCCGAAGGAGCCCAATTTGGGTGCGTAAGTTTGTGACAGTCCC	1425
DB	1201	AA	CTTGTGCTGAACCCCGAAGGAGCCCAATTTGGGTGCGTAAGTTTGTGACAGTCCC	1260
QY	1426	AT	CATCTGACATACCAAGGACAGCTTTTACAAACAGCCCATGCTTACACCTTGGC	1485
DB	1261	AT	CATCTGACATACCAAGGACAGCTTTTACAAACAGCCCATGCTTACACCTTGGC	1320
QY	1486	CAT	TCAGAAGTTCAATCCTGAGGGCTCCACAGAGTGTGGGCTGGTTCAGATCAGAAG	1545
DB	1321	CAT	TCAGAAGTTCAATCCTGAGGGCTCCACAGAGTGTGGGCTGGTTCAGATCAGAAG	1380
QY	1546	AAC	GACCTGGACGACGTGGCAGCTGATGCATCCCGATGGCTCTGCTGTGTGGTCTGCTGA	1605
DB	1381	AAC	GACCTGGACGACGTGGCAGCTGATGCATCCCGATGGCTCTGCTGTGTGGTCTGCTGA	1440

Qy	1606	ACCCGCTCCTCTAAGGATGGCCCTTTACCATCAAGNATCCTGCTGGGCTTCTGTGAG	1665
Db	1441	ACCCGCTCCTCTAAGGATGGCCCTTTACCATCAAGNATCCTGCTGGGCTTCTGTGAG	1500
Qy	1666	ACAATCTCACCTGGCTACTCCATTCACACCTTACCTGTGGGTGCGCAGTGA	1716
Db	1501	ACAATCTCACCTGGCTACTCCATTCACACCTTACCTGTGGGTGCGCAGTGA	1551
RESULT 11			
ID	ABK40235		
ID	ABK40235 standard; cDNA; 1551 BP.		
AC	ABK40235;		
XX			
DT	21-MAY-2002 (first entry)		
DE	DNA encoding human glucocerebrosidase.		
XX			
KW	Human; glucocerebrosidase; GCB; antidiabetic; cytostatic; cancer;		
KW	antiinflammatory; dermatological; immunosuppressive; antirheumatic;		
KW	antiarthritic; antithyroid; detergent; textile; food product;		
KW	agrochemical; feed product; viral pathogen; parasite; autoantigen;		
KW	autoimmune disease; insulin-dependent diabetes mellitus; gene therapy;		
KW	autoimmune thyroid disorder; inflammatory condition; sperm antigen;		
XX	gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200202597-A2.		
XX			
PD	10-JAN-2002.		
XX			
PF	29-JUN-2001; 2001WO-DK00459.		
XX			
PR	30-JUN-2000; 2000DK-0001027.		
PR	14-JUL-2000; 2000DK-0001092.		
PR	29-DEC-2000; 2000WO-DK00743.		
PR	09-FEB-2001; 2001WO-DK00090.		
XX			
PA	(MAXY-) MAXYGEN APS.		
XX			
PA	(MAXY-) MAXYGEN HOLDINGS LTD.		
XX			
PI	Okkels JS, Jensen AD, Van Den Hazel B;		
XX			
DR	WPI; 2002-195745/25.		
DR	P-PSDB: AAU86020.		

CC can be a therapeutic polypeptide useful in human or veterinary therapy, a

CC diagnostic polypeptide used in diagnostics, an industrial polypeptide
 CC useful in the manufacture of products such as detergents, household
 CC articles, textile, food products, agrochemicals, feed products,
 CC polypeptides or proteins associated with viral pathogens, antigens from
 CC parasites, autoantigens associated with autoimmune diseases such as
 CC insulin-dependent diabetes mellitus; autoimmune thyroid disorders;
 CC inflammatory conditions; sperm antigens; and bacterial and cancer
 CC antigens. When the polypeptide is an in vivo glycosylated
 CC polypeptide which does not comprise any other type of non-peptide
 CC moiety, a nucleotide sequence encoding the polypeptide can be used in
 CC gene therapy. When the polypeptide Pp is an antigen, the polypeptide
 CC can be provided in the form of a vaccine. ABK40235-ABK40253 represent
 CC human glucocerebrosidase coding sequences, PCR primers and related
 CC coding sequences of the invention.
 XX
 SQ Sequence 1551 BP; 345 A; 470 C; 389 G; 347 T; 0 other;

Query Match 67.7%; Score 1543; DB 24; Length 1551;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCTCACAGGATTGCTTACTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 225
 DB 1 ATGGCTGGCAGCTCACAGGATTGCTTACTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 60
 QY 226 CGCCCTCGATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 285
 DB 61 CGCCCTCGATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGCTGCAATGCCACA 120
 QY 286 TACTGTGACTCTTTGACCCCGACCTTCTGCTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 345
 DB 121 TACTGTGACTCTTTGACCCCGACCTTCTGCTTTCAGGCAGTGTGCTGGGCATCAGGTGCC 180
 QY 346 AGTACACGAGTGGCGAGCTGAGTGTGATGGGCGCCATCCAGGCTTAATCACACG 405
 DB 181 AGTACACGAGTGGCGAGCTGAGTGTGATGGGCGCCATCCAGGCTTAATCACACG 240
 QY 406 GGCACAGGCTGCTACTGACCTCGAGCCGACAGCAAGATGTCAGAAAGTGAAGGATTT 465
 DB 241 GGCACAGGCTGCTACTGACCTCGAGCCGACAGCAAGATGTCAGAAAGTGAAGGATTT 300
 QY 466 GGAGGGGCGATGACAGATGCTGCTCTCAACATCTTGCCTGTACCCCTGCCCAA 525
 DB 301 GGAGGGGCGATGACAGATGCTGCTCTCAACATCTTGCCTGTACCCCTGCCCAA 360
 QY 526 AATTTGCTACTTAATCTGCTCTCTCTGAAGAAGATCGGATATACATCATCCGGGTA 585
 DB 361 AATTTGCTACTTAATCTGCTCTCTCTGAAGAAGATCGGATATACATCATCCGGGTA 420
 QY 586 CCATGGGCGAGTGTGACTTCTCCATCCGACCTACACCTATGACAGACCCCTGATGAT 645
 DB 421 CCATGGGCGAGTGTGACTTCTCCATCCGACCTACACCTATGACAGACCCCTGATGAT 480
 QY 646 TTCCAGTTGCAACATTCAGGCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
 DB 481 TTCCAGTTGCAACATTCAGGCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540
 QY 706 CACCGAGGCTGCTGCTGGCCAGGCTCCGCTTTCACCTTTCAGGCTGACATCA 765
 DB 541 CACCGAGGCTGCTGCTGGCCAGGCTCCGCTTTCACCTTTCAGGCTGACATCA 600
 QY 766 CCACCTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCC 825
 DB 601 CCACCTTGGCTCAAGACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGACAGCC 660
 QY 826 GGAGACATCTACACAGACCTGGGCCAGATATTTGTGAAGTTCCTGGATGCTATGCT 885
 DB 661 GGAGACATCTACACAGACCTGGGCCAGATATTTGTGAAGTTCCTGGATGCTATGCT 720
 QY 886 GAGCACAGTTACAGTTCTGGGCGAGTACAGCTGAAATGAGCCCTTCTGCTGGGCTGTTG 945
 DB 721 GAGCACAGTTACAGTTCTGGGCGAGTACAGCTGAAATGAGCCCTTCTGCTGGGCTGTTG 780

QY 946 AGTGGATACCCCTTCAGTGCCTGGGCTTCAACCCCTGACATCAGGAGACTTCATTGCC 1005
 DB 781 AGTGGATACCCCTTCAGTGCCTGGGCTTCAACCCCTGACATCAGGAGACTTCATTGCC 840
 QY 1006 CBTGACCTAGTCTTACCTCCGCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 1065
 DB 841 CGTGACCTAGTCTTACCTCCGCAACAGTACTCACCACAATGTCCGCTACTCATGCTG 900
 QY 1066 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
 DB 901 GATGACCAACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 QY 1126 GCTAAATATGCTCATGCAATGCTGTACATGTGTACTGAGCTTTCTGGTTCAGCAAA 1185
 DB 961 GCTAAATATGCTCATGCAATGCTGTACATGTGTACTGAGCTTTCTGGCTCCAGCAAA 1020
 QY 1186 GCACCTTAGGGGAGACACACCGCTGTTCCCAACACCATGCTTTCCTCAGAGGCC 1245
 DB 1021 GCACCTTAGGGGAGACACACCGCTGTTCCCAACACCATGCTTTCCTCAGAGGCC 1080
 QY 1246 TGTGTGGGCTCCAAAGTTCCTGGGAGCAGAGTGTGGCTTCCCTGGATCGAGGATG 1305
 DB 1081 TGTGTGGGCTCCAAAGTTCCTGGGAGCAGAGTGTGGCTTCCCTGGATCGAGGATG 1140
 QY 1306 CAGTACAGCACAAGCATATCAGAACCTCTCTGTACATGTGTGCTGGCTGGAGGACTGG 1365
 DB 1141 CAGTACAGCACAAGCATATCAGAACCTCTCTGTACATGTGTGCTGGCTGGAGGACTGG 1200
 QY 1366 AACCTTGGCTGAAACCCGAGGAGGACCAATTTGGTGGCTAACTTGTGACAGTCCC 1425
 DB 1201 AACCTTGGCTGAAACCCGAGGAGGACCAATTTGGTGGCTAACTTGTGACAGTCCC 1260
 QY 1426 ATCATTTGTAGACATCACCAGGACAGCTTTTACAAACAGCCCATGTTTACACACTTGGC 1485
 DB 1261 ATCATTTGTAGACATCACCAGGACAGCTTTTACAAACAGCCCATGTTTACACACTTGGC 1320
 QY 1486 CACTTACGAAAGTTCATCTCTGAGGCTCCAGAGAGTGGGGCTGTTGCGAGTCAAGAG 1545
 DB 1321 CATTTCAGCAAGTTCATCTCTGAGGCTCCAGAGAGTGGGGCTGTTGCGAGTCAAGAG 1380
 QY 1546 AAGCACTGGACGAGTGGCACTGATGCACTCCGATGGGCTCTGCTGTTGTTGCTGCTA 1605
 DB 1381 AAGCACTGGACGAGTGGCACTGATGCACTCCGATGGGCTCTGCTGTTGTTGCTGCTA 1440
 QY 1606 AAGCGCTCTCTAAGGATGCTCTTACCATCAAGAGTCTGCTGTTGCTGCTGCTGAG 1665
 DB 1441 AAGCGCTCTCTAAGGATGCTCTTACCATCAAGAGTCTGCTGTTGCTGCTGCTGAG 1500
 QY 1666 ACAATCTACCTGGCTACTCCATTCACACTTACCTGCTGGCTGGCCAGTGA 1716
 DB 1501 ACAATCTACCTGGCTACTCCATTCACACTTACCTGCTGGCTGGCCAGTGA 1551

RESULT 12

AAH99924

ID AAH99924 standard; cDNA; 1592 BP.

XX AAH99924;

XX AAH99924;

XX 13-FEB-2002 (first entry)

XX Nucleotide sequence of human glucocerebrosidase.

XX Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
 KW Farber disease; Tay-sachs disease; Niemann-pick disease;
 KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
 KW Gaucher disease; glycosylation; ss.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FH 13..1563
 CDS /*tag= a
 FT

FT XX /product= "Human glucocerebrosidase"
 PN WO200177307-A2.
 PD 18-OCT-2001.
 PF 06-APR-2001; 2001WO-US11144.
 PR 06-APR-2000; 2000US-195598P.
 RA (CYTO-) CYTOLONAL PHARM INC.
 FI Berent SL;
 DR WPI; 2002-041292/05.
 DR P-PSDB; AAG78410.
 XX Novel expression system useful for producing glucocerebrosidase for
 PT treating Gaucher's disease, comprises an insect cell transformed with a
 PT vector encoding glucocerebrosidase that synthesizes glucocerebrosidase
 PT
 PS Claim 2; Page 62-64; 74pp; English.
 XX This invention relates to an expression system comprising an insect
 CC cell transformed with a vector encoding glucocerebrosidase (GC) that
 CC synthesizes clinically effective GC. Recombinant GC is useful for
 CC treating individuals with deficiencies in GC by introducing the
 CC clinically effective form of recombinant GC from insect cells into
 CC individuals. The expression system is stably transformed and
 CC provides a consistently higher level of expression of GC than in
 CC baculovirus or mammalian cell expression systems, and proper
 CC glycosylation modifications for GC, requiring no enzymatic
 CC carbohydrate remodeling to be clinically effective. The expression
 CC system is more effective, economical and simple for the manufacture of
 CC recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack
 CC of lysosomal enzymes contributes to some of the many genetically
 CC inherited lysosomal storage diseases, such as Fabry disease, Farber
 CC disease, Tay-Sachs disease, Niemann-Pick disease, Hunter syndrome,
 CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease.
 CC This nucleotide sequence represents that of human
 CC glucocerebrosidase. See also AAH99925 for the nucleotide sequence of
 CC human glucocerebrosidase, noting the differences at residues
 CC 1552 to 1554. This sequence encodes arginine at that position,
 CC AAH99925 encodes histidine.
 XX
 XX Sequence 1592 BP; 359 A; 481 C; 397 G; 355 T; 0 other;

Query Match 67.1%; Score 1529.6; DB 24; Length 1592;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1541; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 157 GTAAGCATCATGGCGGCGGCGCTACAGGATGCTTCTACTTCAGGCGAGTTCGGGCA 216
 DB 4 GAATTCAAATGGCTGGCTCGTTACAGGATGCTTCTACTTCAGGCGAGTTCGGGCA 63
 QY 217 TCAGGTGCGCGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTTCGGTGTGTCTGTC 276
 DB 64 TCCGGAGCTAGACCTTGCATCCCTAAAAGCTTCGGCTACAGCTTCGGTGTGTCTGTC 123
 QY 277 AATGCCACATCTGTGACTCTCTTTGACCCCGGACCTTTCTCTGCCCTTGGTACCTTCAGC 336
 DB 124 AATGCCACATCTGTGACTCTCTTTGACCCCGGACCTTTCTCTGCCCTTGGTACCTTCAGC 183
 QY 337 CGCTATGAGAGTACACGAGTGGGCGAGGATGAGTATGGGGCCCATCCAGGCT 396
 DB 184 CGCTATGAGAGTACACGAGTGGGCGAGGATGAGTATGGGGCCCATCCAGGCT 243
 QY 397 AATCACGGGCGACAGGCTCTCTACTCACCTTCAGCGACAGAGTTCAGAAAGTG 456
 DB 244 AATCACGGGCGACAGGCTCTCTACTCACCTTCAGCGACAGAGTTCAGAAAGTG 303
 QY 457 AAGGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTGGCTGTGACCC 516

DB 304 AAGGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCCTTGGCTGTGACCC 363
 QY 517 CTTGCCCAAAATTTGCTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAACATC 576
 DB 364 CTTGCCCAAAATTTGCTTAAATCGTACTTCTCTGAAGAGGAATCGGATATAACATC 423
 QY 577 ATCCGGGTACCCATGCGCCAGCTGTGACTTCTCCATCCGACCTACACTATGCGAGACCC 636
 DB 424 ATCCGGGTACCCATGCGCCAGCTGTGACTTCTCCATCCGACCTACACTATGCGAGACCC 483
 QY 637 CTTGATGATTTCCAGTTGACAACTTCAAGCTTCCGAGAGGAATACCAAGCTCAAGATA 696
 DB 484 CTTGATGATTTCCAGTTGACAACTTCAAGCTTCCGAGAGGAATACCAAGCTCAAGATA 543
 QY 697 CCGCTGATTTACCGAGCGCTGCAAGTTGGCCCAAGCTTCCCTTCTCACTTCCGAGCCCC 756
 DB 544 CCGCTGATTTACCGAGCGCTGCAAGTTGGCCCAAGCTTCCCTTCTCACTTCCGAGCCCC 603
 QY 757 TGGACATCACCCACTTGGCTCAAGCAATGAGCGGTGAATGGGAAGGGCTCAGTCAG 816
 DB 604 TGGACATCACCCACTTGGCTCAAGCAATGAGCGGTGAATGGGAAGGGCTCAGTCAG 663
 QY 817 GGACAGCCCGGAGACATCTACCAACAGAGCTGGGCCAGTACTTTGTGAAGTTCTTGGAT 876
 DB 664 GGACAGCCCGGAGACATCTACCAACAGAGCTGGGCCAGTACTTTGTGAAGTTCTTGGAT 723
 QY 877 GCGTATGCTGAGCACAAGTTTACAGTTCTGGGAGTGAGAGCTGAAATGAGCTTCTGCT 936
 DB 724 GCGTATGCTGAGCACAAGTTTACAGTTCTGGGAGTGAGAGCTGAAATGAGCTTCTGCT 783
 QY 937 GGGCTGTGAGTGAGTACCCCTTCCAGTGGCTTCCAGCTGGGCTTCAACATCAGCGAGAC 996
 DB 784 GGGCTGTGAGTGAGTACCCCTTCCAGTGGCTTCCAGCTGGGCTTCAACATCAGCGAGAC 843
 QY 997 TTCAATCCCGCTGACCTAGTGTCTACCTCGCCCAACAGTACTCACCACATGTCGCGCTA 1056
 DB 844 TTCAATCCCGCTGACCTAGTGTCTACCTCGCCCAACAGTACTCACCACATGTCGCGCTA 903
 QY 1057 CTCATGCTGATGATCACAAGCTTGTGCTGTCGCCACAGTGGGCAAGTGGTACTGACAGAC 1116
 DB 904 CTCATGCTGATGATCACAAGCTTGTGCTGTCGCCACAGTGGGCAAGTGGTACTGACAGAC 963
 QY 1117 CCAGAAGCAGCTAAATATGTTCTATGCGATTCGCTGATGCTGATGCTGATGCTGCTGCT 1176
 DB 964 CCAGAAGCAGCTAAATATGTTCTATGCGATTCGCTGATGCTGATGCTGATGCTGCTGCT 1023
 QY 1177 CCAGCCAAAGCCACCTAGGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTGGC 1236
 DB 1024 CCAGCCAAAGCCACCTAGGGAGACACACCGCTGTTCCTCCCAACACCATGCTCTTGGC 1083
 QY 1237 TCAGAGCCCTGTGCGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGAT 1296
 DB 1084 TCAGAGCCCTGTGCGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGAT 1143
 QY 1297 CGAGGGATGCAAGTACAGCCACAGCATCATCAGAACCTCTGTACCATGTGTCGCGTGG 1356
 DB 1144 CGAGGGATGCAAGTACAGCCACAGCATCATCAGAACCTCTGTACCATGTGTCGCGTGG 1203
 QY 1357 ACCGACTTGAACCTTCCCTGAAACCCGAGGAGGCCAAATTTGGGTGCGTAACTTGTG 1416
 DB 1204 ACCGACTTGAACCTTCCCTGAAACCCGAGGAGGCCAAATTTGGGTGCGTAACTTGTG 1263
 QY 1417 GACAGTCCCATCATGTTGAGATCAGCAAGGAGACAGTGTTCACAAACAGCCCATGTTCTAC 1476
 DB 1264 GACAGTCCCATCATGTTGAGATCAGCAAGGAGACAGTGTTCACAAACAGCCCATGTTCTAC 1323
 QY 1477 CACCTTGGGCGCTTACAGCAAGTGTTCCTTGGGCTCCAGAGAGTGGGGTGGTGGC 1536
 DB 1324 CACCTTGGGCGCTTACAGCAAGTGTTCCTTGGGCTCCAGAGAGTGGGGTGGTGGC 1383
 QY 1537 AGTCAGAAGAACGACCTGGAGCAGTGGCACTGATGATCCCGATGGCTCTCTGTTGTG 1596

```
Db 1384 AGTCAGAGAGAGAGCTGCGACGACGAGTGGCAGCTGATGCATCCGATGGCTGCTGCTG 1443
QY 1597 GTCGTGCTAAACGGCTCCTCTAAGATGTCCTCTTACATCAAGGATCCCTGCTGTGGC 1656
Db 1444 GTCGTGCTAAACGGCTCCTCTAAGATGTCCTCTTACATCAAGGATCCCTGCTGTGGC 1503
QY 1657 TTCTGCGAGACAATCTCACTGGCTACTGCTACCTCAATTCACACTACCTGCTGGCGTGCAGTGA 1716
Db 1504 TTCTGCGAGACAATCTCACTGGCTACTGCTACCTCAATTCACACTACCTGCTGGCGTGCAGTGA 1563

RESULT 13
AAQ39286
ID AAQ39286 standard; DNA; 7620 BP.
XX
AC AAQ39286;
XX
DT 20-JUL-1993 (first entry)
XX
DE Glucocerebrosidase gene.
XX
KW Glucocerebrosidase; peripheral blood leukocyte; lysosomal degradation;
KW glycolipid; Gaucher disease; glucosylceramide; glucocerebrosidase; RFLP;
KW restriction fragment length polymorphism; mutation; pseudogene; 1226G;
KW Jewish; 1448C; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 356..611
FT /*tag= a
FT /number= Exon 1
FT Intron 612..979
FT /*tag= b
FT /number= Intron 1
FT exon 980..1067
FT /*tag= c
FT /number= Exon 2
FT Intron 1068..1619
FT /*tag= d
FT /number= Intron 2
FT exon 1620..1811
FT /*tag= e
FT /number= Exon 3
FT Intron 1812..1934
FT /*tag= f
FT /number= Intron 3
FT exon 1935..2081
FT /*tag= g
FT /number= Exon 4
FT Intron 2082..3046
FT /*tag= h
FT /number= Intron 4
FT exon 3047..3180
FT /*tag= i
FT /number= Exon 5
FT Intron 3181..3390
FT /*tag= j
FT /number= Intron 5
FT exon 3391..3563
FT /*tag= k
FT /number= Exon 6
FT Intron 3564..4116
FT /*tag= l
FT /number= Intron 6
FT exon 4117..4354
FT /*tag= m
FT /number= Exon 7
FT Intron 4355..5227
FT /*tag= n
FT /number= Intron 7
FT exon 5228..5452
FT /*tag= o
```

```
FT intron /number= Exon 8
FT 5453..5852 /*tag= p
FT /number= Intron 8
FT 5853..6016
FT /*tag= q
FT /number= Exon 9
FT 6017..6385 /*tag= r
FT /number= Intron 9
FT 6386..6502 /*tag= s
FT /number= Exon 10
FT 6503..6596 /*tag= t
FT /number= Intron 10
FT 6597..7245 /*tag= u
FT /number= Exon 11
FT
FT
XX WO9306244-A.
XX
XX 01-APR-1993.
XX
XX 16-SEP-1992; 92WO-US07840.
XX
XX 27-SEP-1991; 91US-0767135.
XX
XX (SCRI ) SCRIPPS RES INST.
XX
XX Beutler E, Sorge JA;
XX
XX WPI; 1993-117560/14.
XX
XX Screening method for new Gaucher disease mutation - comprises
XX amplifying DNA by PCR and assaying the prod. with probe to
XX guanine insertion adjacent position 57 of gluco-cerebrosidase gene-
XX exon 2
XX
XX Disclosure; Page 55-60; 74pp; English.
XX
XX This sequence represents the glucocerebrosidase gene which was
XX isolated from peripheral blood leukocytes. Glucocerebrosidase is an
XX enzyme which is required for the lysosomal degradation of glycolipids.
XX A deficiency of this enzyme leads to Gaucher disease, as in the absence
XX of glucocerebrosidase, the extremely insoluble glucosylceramide
XX (glucocerebroside) accumulates. The gene for glucocerebrosidase is
XX located on chromosome 1, q21. A number of different mutations cause
XX Gaucher disease and analysis of these mutations has been complicated
XX by the presence of a pseudogene 16 kb downstream from the glucocere-
XX brosidase gene. The pseudogene is approx. 95% homologous to the
XX functional gene and can be transcribed but not translated. About
XX 75% of Jewish patients with clinically significant Gaucher disease
XX contain a characteristic A>G mutation at cDNA nucleotide 1226 (1226G
XX mutation) which corresponds to amino acid 370 of the mature protein.
XX The corresponding position of the mutation in the glucocerebrosidase
XX gene is in exon 9, position 2. The same mutation is found in approx.
XX 25% of non-Jewish disease causing alleles. This mutation causes a
XX characteristic restriction fragment length polymorphism (RFLP). A
XX second, less common mutation is at cDNA position 1448 and is a T>C
XX mutation (1448C mutation). The corresponding position of this
XX mutation is exon 10, position 60. The 1448C mutation only accounts
XX for 2% of Jewish Gaucher disease producing alleles and for about 40%
XX of the alleles in non-Jewish patients.
XX
XX Sequence 7620 BP; 1665 A; 2091 C; 1925 G; 1939 T; 0 other;
```

Query Match 29.3%; Score 667; DB 14; Length 7620;

Best Local Similarity 89.1%; Pred. No. 1.le-175;

Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;

QY 1489 TTCAGCAAGTTTCATTCCTGAGGGCTCCAGAGAGTGGCGCTGGTTCAGAGTCAAGAAC 1548

|||||

Db 6381 TTCAGCAAGTTCTCTGAGGCTCCAGAGAGTGGGCTTGGCCAGTCAAGAAGAC 6440
 QY 1549 GACCTGACCACTGGCAGTGAATCCGATGGCTCTCTGTTGTGCTGCTGAAC 1608
 Db 6441 GACCTGACCACTGGCAGTGAATCCGATGGCTCTCTGTTGTGCTGCTGAAC 6500
 QY 1609 C----- 1609
 Db 6501 CGGTGAGGCAATGGTGAAGTCTGGGCTGAAGACAGCGTTGGGGCCCTTGCA 6560
 QY 1610 -----GCTCTCTAAGGATGTGCCTCTTAC 1634
 Db 6561 GGATCACACTCTCAGCTTCTCTCCCTGCTCCCTAGCTCTCTAAGGATGTGCCTTAC 6620
 QY 1635 CATCAAGGATCTCTGTTGGGCTTCTCTGGAGACAATCTCACTGGCTACTTCCATTCAAC 1694
 Db 6621 CATCAAGGATCTCTGTTGGGCTTCTCTGGAGACAATCTCACTGGCTACTTCCATTCAAC 6680
 QY 1695 CTACTGTGGCTGCGCAGTGTGAGCAGATCTCAAGGAGCACTGGGCTCAGCCTGG 1754
 Db 6681 CTACTGTGGCTGCGCAGTGTGAGCAGATCTCAAGGAGCACTGGGCTCAGCCTGG 6740
 QY 1755 GCATTAAGGAGCAGATGTCAGCTCACAGCTGTCTGTGACTAAGAGGCGCAGCAGGC 1814
 Db 6741 GCATTAAGGAGCAGATGTCAGCTCACAGCTGTCTGTGACTAAGAGGCGCAGCAGGC 6800
 QY 1815 CAGGTGAGCTTACAGCAGCTTAAGCCAGGCGCAATGGTTGGTGTACTTCTTCCCT 1874
 Db 6801 CAGGTGAGCTTACAGCAGCTTAAGCCAGGCGCAATGGTTGGTGTACTTCTTCCCT 6860
 QY 1875 TCTAGTGTGCGCAGGCTGAGGCGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCTCC 1934
 Db 6861 TCTAGTGTGCGCAGGCTGAGGCGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCTCC 6920
 QY 1935 CCAGCCCACTGATGTATGTGAACATGCGCTGTGTGCTGCTTTGGAACTGGGCGCTG 1994
 Db 6921 CCAGCCCACTGATGTATGTGAACATGCGCTGTGTGCTGCTTTGGAACTGGGCGCTG 6980
 QY 1995 GGTCCAGGCTAGGTGAGCTCAGTGTCCGTACAAACACAGATCAGGCTCAGGCTGAAG 2054
 Db 6981 GGTCCAGGCTAGGTGAGCTCAGTGTCCGTACAAACACAGATCAGGCTCAGGCTGAAG 7040
 QY 2055 GAAAGAGAGACTAGGAAGCTGGGCGCAAACTGAGAGCTGTTGTCTTCTTCTGGAGA 2114
 Db 7041 GAAAGAGAGACTAGGAAGCTGGGCGCAAACTGAGAGCTGTTGTCTTCTTCTGGAGA 7100
 QY 2115 TGCAGAACTGGGCGCTGGAGCAGCTGTCAGCATCAGGCGGGAAGCTTAAAGCAGCA 2174
 Db 7101 TGCAGAACTGGGCGCTGGAGCAGCTGTCAGCATCAGGCGGGAAGCTTAAAGCAGCA 7160
 QY 2175 GCGGCTGTGCCAGGCGCAGATGATGCTATGGCAACAGCAGGAGAAATGGCAGCTC 2234
 Db 7161 GCGGCTGTGCCAGGCGCAGATGATGCTATGGCAACAGCAGGAGAAATGGCAGCTC 7220
 QY 2235 TTAAGGAGAAATGTTGAGCCCA 2259
 Db 7221 TTAAGGAGAAATGTTGAGCCCA 7245

RESULT 14

ABL67250
 ID ABL67250 standard; DNA; 5769 BP.

AC ABL67250;
 XX

DT 15-MAY-2002 (first entry)
 XX

DE Thyroid cancer related gene sequence SEQ ID NO:5587.
 XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.

XX Homo sapiens.
 OS WO200194629-A2.
 PN 13-DEC-2001.
 PD 30-MAY-2001; 2001WO-US10838.
 PE 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 PA (AVAL-) AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
 chemical agent to be tested for anti-neoplastic activity, and
 determining a change in expression of a gene of a signature gene set

Claim 1; SEQ ID 5587; 44pp; English.

The present invention describes a method (M1) for screening for an
 anti-neoplastic agent. The method involves exposing cells to a chemical
 agent to be tested for anti-neoplastic activity, determining a change in
 expression of at least one gene (I) of a signature gene set, where (I)
 comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 to ABL70110), or is at least 95% identical to (S), where a change in
 expression is indicative of anti-neoplastic activity. (I) has cytostatic
 activity and can be used in gene therapy. M1 can be used for screening

100

100

100

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 10:29:56 ; Search time 2925 Seconds
(without alignments)
12618.637 Million cell updates/sec

Title: US-10-024-197-24

Perfect score: 2279

Sequence: 1 agctaagcagctacgtca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_man: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1361.8	59.8	1931	11	BC030240
2	921.8	40.4	953	9	AL541128
3	901.6	39.6	915	9	AL571401
4	896.2	39.3	927	9	AL568749
5	879.8	38.6	940	9	AL524792
6	864.6	37.9	920	14	BQ676149

7	858.4	37.7	916	12	BG744064
8	856.8	37.6	957	12	BG744733
9	850.2	37.3	909	14	AL562392
10	841	36.9	931	14	BQ680411
11	835	36.6	838	14	BQ681028
12	833.4	36.6	916	14	BQ677036
13	830.8	36.5	1017	13	BM477383
14	829	36.4	896	14	BQ897517
15	826	36.2	1008	14	BQ069685
16	824	36.2	938	14	BQ683279
17	812.6	35.7	872	14	BQ678499
18	811.8	35.6	896	14	BQ678522
19	811.2	35.6	863	12	BQ281198
20	811	35.6	896	14	BQ683460
21	805.4	35.3	849	9	AL562243
22	801	35.1	954	9	AL523699
23	794.8	34.9	945	12	BQ674486
24	791.4	34.7	923	13	BQ680595
25	779.8	34.2	827	13	B1917380
26	766.6	33.6	861	12	BQ678530
27	763.2	33.5	1043	14	BQ071624
28	754.8	33.1	793	13	B1910946
29	748.6	32.8	775	14	BQ008716
30	748.4	32.8	774	14	BQ574459
31	738	32.4	787	9	AU122294
32	735	32.3	862	13	B1756617
33	733	32.3	913	12	BG765009
34	734.8	32.2	855	13	B1769458
35	731.8	32.1	1070	13	BM544110
36	729.2	32.0	763	14	BQ000311
37	725.8	31.8	795	9	AL572439
38	722.6	31.7	873	14	BQ425733
39	722.4	31.7	725	14	BM772392
40	719.2	31.6	859	12	BG764893
41	715.2	31.4	756	10	B6613411
42	713.4	31.3	773	9	AU138949
43	713	31.3	885	12	BG749231
44	710.6	31.2	814	9	AU140959
45	707.4	31.0	885	12	BE867971

ALIGNMENTS

BC030240 1931 bp mRNA linear HTC 20-MAY-2002
Homo sapiens, Similar to glucosidase, beta; acid (includes
glucosylceramidase), clone IMAGE:5114119, mRNA.
BC030240
BC030240.1 GI:20988849
HTC.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1931)
Strausberg, R.
Direct Submission
Submitted (07-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cdNA Library Preparation: Life Technologies, Inc.
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Db 1373 TGGAGCAGATACCAAGGAGGCACTGGGCTCAGCCTGGGCAATTAAGGAGACAGAGTCAGC 1432
 QY 1777 TCACACGCTGTGTGATTAAGAGGGACACAGCAGGCGAGTGTAGCTTACAGCAGCT 1836
 Db 1433 TCACACGCTGTGTGATTAAGAGGGACACAGCAGGCGAGTGTAGCTTACAGCAGCT 1492
 QY 1837 AAGCCAGGGGCAATGGTTGGTGACTACTTCCCTCTAGTGTGGCAGGGCTGG 1896
 Db 1493 AAGCCAGGGGCAATGGTTGGTGACTACTTCCCTCTAGTGTGGCAGGGCTGG 1552
 QY 1897 AGGCCCTAGAAAAGATCAGTAAGCCAGTGTCCGCCAGCCGCCCATCTTATGTGAA 1956
 Db 1553 AGGCCCTAGAAAAGATCAGTAAGCCAGTGTCCGCCAGCCGCCCATCTTATGTGAA 1612
 QY 1937 CATGCGTGTGTGCTGCTTGTGAACTGGGCTGGGCTAGGCTAGGCTAGGCTC 2016
 Db 1613 CATGCGTGTGTGCTGCTTGTGAACTGGGCTGGGCTAGGCTAGGCTAGGCTC 1672
 QY 2017 ACTGCTGCTACAAACACAGATCAGGCTGAGGCTAAGGAAAGAGAGACTAGGAAAGC 2076
 Db 1673 ACTGCTGCTACAAACACAGATCAGGCTGAGGCTAAGGAAAGAGAGACTAGGAAAGC 1732
 QY 2077 TGGGCCCAAACTGGAGACTGTTTCTTCTGAGATGACAGACTGGGCCGCTGGAGC 2136
 Db 1733 TGGGCCCAAACTGGAGACTGTTTCTTCTGAGATGACAGACTGGGCCGCTGGAGC 1792
 QY 2137 AGCAGTGTGAGATCAGGCGGAGGCTTAAAGCAGCAGGCGGTGGCCAGCCAGC 2196
 Db 1793 AGCAGTGTGAGATCAGGCGGAGGCTTAAAGCAGCAGGCGGTGGCCAGCCAGC 1852
 QY 2197 ATGATTCCTATGCGACCCAGGAGAAATGGCAGCTTCTTAAAGGAGAGAAATGTTTGGC 2256
 Db 1853 ATGATTCCTATGCGACCCAGGAGAAATGGCAGCTTCTTAAAGGAGAGAAATGTTTGGC 1912
 QY 2257 CCAAAAAAATAAAAAA 2275
 Db 1913 CCAAAAAAATAAAAAA 1931

AL541128 953 bp mRNA linear EST 16-FEB-2001
 AL541128 LTI_FL002_P1 Homo sapiens cDNA clone CS0DE005YL08 5 prime
 / mRNA sequence.
 AL541128
 AL541128.1 GI:12871896
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 L1, W.B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..953
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE005YL08"
 /clone_lib="LTI_FL002_P1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville

BASE COUNT 214 a 271 c 240 g 223 t 5 others
 ORIGIN
 Query Match 40.4%; Score 921.8; DB 9; Length 953;
 Best Local Similarity 99.1%; Pred. No. 5e-172;
 Matches 942; Conservative 5; Mismatches 2; Indels 2; Gaps 2;
 QY 1 AGCTAAGCAGGATACCTGTCATCCTCTGTTTGTAGTGGATCCTCTATCTTCCAGAGAC 60
 Db 4 AGCTAAGCAGGATACCTGTCATCCTCTGTTTGTAGTGGATCCTCTATCTTCCAGAGAC 63
 QY 61 TCTGAACCCCTGTGGTCTTCTTCTCATCTAATGACCTGAGGGATGGAGTTTCAAGT 120
 Db 64 TCTGAACCCCTGTGGTCTTCTTCTCATCTAATGACCTGAGGGATGGAGTTTCAAGT 123
 QY 121 CTTTCCAGAGAGGATGTCCTCAAGCCTTTGAGTGGTAAAGCATCATGCTGCGCAGCCTC 180
 Db 124 CTTTCCAGAGAGGATGTCCTCAAGCCTTTGAGTGGTAAAGCATCATGCTGCGCAGCCTC 183
 QY 181 ACAGAGTGTCTTACTTTCAGGAGTGTCTGGGATCAGTGGCCGCCCTGCATCCCT 240
 Db 184 ACAGAGTGTCTTACTTTCAGGAGTGTCTGGGATCAGTGGCCGCCCTGCATCCCT 243
 QY 241 AAAGCTTCGGCTACAGCTCGGTGGTGTCTGTCTCAATGCCACATATCTGTGACTCCTTT 300
 Db 244 AAAGCTTCGGCTACAGCTCGGTGGTGTCTGTCTCAATGCCACATATCTGTGACTCCTTT 303
 QY 301 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGCTGGG 360
 Db 304 GACCCCGGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGAGCTGGG 363
 QY 361 CGAGGATGGAGTGTAGTATGGGGCCATCCAGGCTAATCACGGGCACAGGCTGTGTA 420
 Db 364 CGAGGATGGAGTGTAGTATGGGGCCATCCAGGCTAATCACGGGCACAGGCTGTGTA 423
 QY 421 CTGACCTTCGAGCAGACAGAGTTCAGAGAGTGAAGGATTTGGAGGGCCATGACA 480
 Db 424 CTGACCTTCGAGCAGACAGAGTTCAGAGAGTGAAGGATTTGGAGGGCCATGACA 483
 QY 481 GATGCTGCTCTCTCAACATCTTCCCTGTGTCACCCCTGCCAAAATTTGCTACTTAAA 540
 Db 484 GATGCTGCTCTCTCAACATCTTCCCTGTGTCACCCCTGCCAAAATTTGCTACTTAAA 543
 QY 541 TCGTACTTCTCTGAAGAGATFCGGATATACATCATFCGGGTACCCATGGCCAGCTGT 600
 Db 544 TCGTACTTCTCTGAAGAGATFCGGATATACATCATFCGGGTACCCATGGCCAGCTGT 603
 QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCAAC 660
 Db 604 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCAAC 663
 QY 661 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGTACCCCTGATTTCCAGGAGCCCTGAG 720
 Db 664 TTCAGCCTCCAGAGGAGATACCAAGCTCAAGTACCCCTGATTTCCAGGAGCCCTGAG 723
 QY 721 TTGGCCAGCCTGCTTCTTCTCTCTTCCAGCCCTGAGATACCCCTGATTTCCAGGAGCCCTGAG 780
 Db 724 TTGGCCAGCCTGCTTCTTCTCTCTTCCAGCCCTGAGATACCCCTGATTTCCAGGAGCCCTGAG 783
 QY 781 ACCAATGAGCGTGAATGGAGGGTCACTCAAGGAGAGCCGCGAGACATCTACAC 840
 Db 784 ACCAATGAGCGTGAATGGAGGGTCACTCAAGGAGAGCCGCGAGACATCTACACMA - 842
 QY 841 CAGACCTTGGCCAGATACCTTGTGAAGTTCCTGGATGCTGATGCTGAGCAGCAAGTTACAG 900
 Db 843 CAGACCTTGGCCAGATACCTTGTGAAGTTCCTGGATGCTGATGCTGAGCAGCAAGTTACAG 902
 QY 901 TTCTGGCAGTACAGCTG - AAAATGAGCCCTTCTGCTGGGCTTGTGAGTGG 950
 Db 903 TTCTGGGAGTACAGCTGAAAAATGAGCCCTTCTGCTGGGCKGTTGAGTGG 953

```
RESULT 3
AL571401/c
LOCUS          915 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION    AL571401 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1023Y009 3
prime, mRNA sequence.
ACCESSION     AL571401
VERSION       AL571401.1 GI:12928661
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 915)
AUTHORS       Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE         Full-length cDNA libraries and normalization
JOURNAL       Unpublished (2001)
COMMENT       Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
               Location/Qualifiers
                 i. 915
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="CS0D1023Y009"
                   /clone_lib="LTI_NFL006_PL2"
                   /tissue="placenta"
                   /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT    208 a 234 c 270 g 201 t      2 others
ORIGIN
Query Match   39.6%; Score 901.6; DB 9; Length 915;
Best Local Similarity 99.9%; Pred. No. 4.9e-168;
Matches 901; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 835 TACCACGACCTGGCCAGATCTTGTGAAGTTCTCTGATGCTCTATGCTGAGCACAAG 894
DB 915 TACCACGACCTGGCCAGATCTTGTGAAGTTCTCTGATGCTCTATGCTGAGCACAAG 856
QY 895 TTACAGTTCTGGCAGTGACAGCTGAAATGAGCCCTTCTGCTGGGCTGTTGAGTGATAC 954
DB 855 TTACAGTTCTGGCAGTGACAGCTGAAATGAGCCCTTCTGCTGGGCTGTTGAGTGATAC 796
QY 955 CCCTTCAGTGGCTGGCTTACCCCTGAAATGACGAGACTTCATTCGCCGTGACCTA 1014
DB 795 CCCTTCAGTGGCTGGCTTACCCCTGAAATGACGAGACTTCATTCGCCGTGACCTA 736
QY 1015 GGTCTTACCTCCCAACAGTACTCACCAGTGTCCGCTCTATGCTGGATGACCAA 1074
DB 735 GTTCCTACCTCCCAACAGTACTCACCAGTGTCCGCTCTATGCTGGATGACCAA 676
QY 1075 CGCTTGCTGCTGCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAAATAT 1134
DB 675 CGCTTGCTGCTGCCCACTGGGCAAGGTGGTACTGACAGACCCAGAGCAGCTAAATAT 616
QY 1135 GTTCATGGCAATGCTGTACATTTGGTACCTGGAGCTTTCTGGCTCCAGCCAAAGCCACCTA 1194
DB 615 GTTCATGGCAATGCTGTACATTTGGTACCTGGAGCTTTCTGGCTCCAGCCAAAGCCACCTA 556
QY 1195 GGGGAGACACACCGCTGTGCCCAACACCATGCTCTTTGCCCTCAGAGGCGCTGTGGGCG 1254
DB 555 GGGGAGACACACCGCTGTGCCCAACACCATGCTCTTTGCCCTCAGAGGCGCTGTGGGCG 496
```

```
QY 1255 TCCAAGTTCTGGGAGCAGAGTGTGGCTTAGGCTCTCTGGATCGAGGATGCGAGTACAGC 1314
DB 495 TCCAAGTTCTGGGAGCAGAGTGTGGCTTAGGCTCTCTGGATCGAGGATGCGAGTACAGC 436
QY 1315 CACAGCATCATCAGAACCTCTCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGCC 1374
DB 435 CACAGCATCATCAGAACCTCTCTGTACCATGTGTGGCTGGACCGACTGGAACCTTGCC 376
QY 1375 CTGAACCCCGAAGGAGGACCAATTTGGTGGTAACTTTGTCGACAGTCCCATCATTTGA 1434
DB 375 CTGAACCCCGAAGGAGGACCAATTTGGTGGTAACTTTGTCGACAGTCCCATCATTTGA 316
QY 1435 GACATCATCAAGGACAGCTTTTACAAACAGCCCATGTTTACCACTTGGCCACTTCAGC 1494
DB 315 GACATCATCAAGGACAGCTTTTACAAACAGCCCATGTTTACCACTTGGCCACTTCAGC 256
QY 1495 AAGTTTCAATCTCTGAGGGTCTCCAGAGAGTGGGGCTGTTGCCAGTCAAGAGACGACTG 1554
DB 255 AAGTTTCAATCTCTGAGGGTCTCCAGAGAGTGGGGCTGTTGCCAGTCAAGAGACGACTG 196
QY 1555 GACGAGTGGCAGTGTGATGATCCGATGCTCTGCTGTTGTTGGTCTGCTAAACCGCTCC 1614
DB 195 GACGAGTGGCAGTGTGATGATCCGATGCTCTGCTGTTGTTGGTCTGCTAAACCGCTCC 136
QY 1615 TCTAAGGATGCTCTTACCATCAAGGATCCTCTGCTGTGGGCTTCTTGGAGACAACTCA 1674
DB 135 TCTAAGGATGCTCTTACCATCAAGGATCCTCTGCTGTGGGCTTCTTGGAGACAACTCA 76
QY 1675 CCTGGCTACTCCATTCACACCTACTCTGTCGCTGCGCAGTGTGAGCAGATCTCAAGG 1734
DB 75 CCTGGCTACTCCATTCACACCTACTCTGTCGCTGCGCAGTGTGAGCAGATCTCAAGG 16
QY 1735 AG 1736
DB 15 AG 14
```

```
RESULT 4
AL568749/c
LOCUS          927 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION    AL568749 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YL08 3 prime
, mRNA sequence.
ACCESSION     AL568749
VERSION       AL568749.1 GI:12923399
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 927)
AUTHORS       Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE         Full-length cDNA libraries and normalization
JOURNAL       Unpublished (2001)
COMMENT       Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
               Location/Qualifiers
                 i. 927
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="CS0DE005YL08"
                   /clone_lib="LTI_FL002_PL1"
                   /lab_host="DH10B"
                   /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
```


Db 386 CGAGGATGAGTGTGATGGGCGCCATCCAGGCTAATACACGGGCGACAGGCGCTGCTA 445
 QY 421 CTGACCTCCAGCAGACAGAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACA 480
 Db 446 CTGACCTCCAGCAGACAGAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACA 505
 QY 481 GATGCTGCTGCTCAACATCTTGCCTGTCCACCCCTGCGCCCAAAATTTGCTACTTAA 540
 Db 506 GATGCTGCTGCTCAACATCTTGCCTGTCCACCCCTGCGCCCAAAATTTGCTACTTAA 565
 QY 541 TCGTACTTCTCTGAAGAGGATTCGATATAACATATCCCGGTACCCATGGCCAGCTGT 600
 Db 566 TCGTACTTCTCTGAAGAGGATTCGATATAACATATCCCGGTACCCATGGCCAGCTGT 625
 QY 601 GACTTCTCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 660
 Db 626 GACTTCTCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 685
 QY 661 TTGAGCTCCAGAGAGGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
 Db 686 TTGAGCTCCAGAGAGGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 745
 QY 721 TTGGCCAGGCTCCCTTTTCACTCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAG 780
 Db 746 TTGGCCAGGCTCCCTTTTCACTCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAG 805
 QY 781 ACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGAGACCCCGGAGACATCTACAC 840
 Db 806 ACCAATGGAGCGGTGAATGGAGGGGTCACTCAAGGGAGACCCCGGAGACATCTACAC 864
 QY 841 CAGACCTGGGCCAGATPACTTTGTGAAGTTCCTGGATGCTTATGCTGAGCACAAGTTACAG 900
 Db 865 CAGACCTGGGCCAGATPACTTTGTGAAGTTCCTGGATGCTTATGCTGAGCACAAGTTACAG 922
 QY 901 TTCTGGGCGAGTGACAGC 917
 Db 923 TTCTGGGCGAGTGACAGC 939

RESULT 6
 BQ676149
 LOCUS BQ676149 920 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8494833 NIH_MGC_112 Homo sapiens cdna clone IMAGE:6294994
 5', mRNA sequence.
 ACCESSION BQ676149
 VERSION BQ676149.1 GI:21788828
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 920)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 cdna Library Preparation: Rubin Laboratory
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DATA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLM2500 row: i column: 11
 High quality sequence start: 19
 High quality sequence stop: 704.
 Location/Qualifiers

FEATURES
 source
 1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6294994"
 /clone_lib="NIH_MGC_112"

/tissue_type="melanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site: 2:
 EcoRI; cdna made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

BASE COUNT 214 a 279 c 225 g 202 t

ORIGIN

Query Match 37.9%; Score 864.6; DB 14; Length 920;
 Best Local Similarity 98.2%; Pred. No. 9.8e-161;
 Matches 895; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
 QY 261 GGTGGTGTGTCTGCAATGCCACATACTGTGACTCTTTGACCCCGGACCTTTCCTGC 320
 Db 10 GGTGGTGTGTCTGCAATGCCAC-TACTGTGACTCTTTGACCCCGGACCTTTCCTGC 68
 QY 321 CCTGGTACCTTTCAGCGCTATGAGGTACACGAGTGGGCGAGGATGGAGCTGAGTAT 380
 Db 69 CCTGGTACCTTTCAGCGCTATGAGGTACACGAGTGGGCGAGGATGGAGCTGAGTAT 128
 QY 381 GGGGCCCATCCAGGCTAATCACACGGGCGACAGGCTGCTACTGACCTGCGAGCAGAAC 440
 Db 129 GGGGCCCATCCAGGCTAATCACACGGGCGACAGGCTGCTACTGACCTGCGAGCAGAAC 188
 QY 441 GAACTTCAGAAAGTGAAGGATTTGGAGGGGCGCATGACAGATGCTGCTCTCAACAT 500
 Db 189 GAACTTCAGAAAGTGAAGGATTTGGAGGGGCGCATGACAGATGCTGCTCTCAACAT 248
 QY 501 CCTTGGCTGTCAACCCCTGCGCCAAATTTCTACTTAAATCGTACTTCTCTGAAGAAG 560
 Db 249 CCTTGGCTGTCAACCCCTGCGCCAAATTTCTACTTAAATCGTACTTCTCTGAAGAAG 308
 QY 561 AATGGATATACATCATCCGGGTACCATGGGCGAGCTGTGACTTCTCCATCCGACCTA 620
 Db 309 AATGGATATACATCATCCGGGTACCATGGGCGAGCTGTGACTTCTCCATCCGACCTA 368
 QY 621 CACCTATCAGACACCCCTGTATGATTTCCAGTTGACAACTTCCAGCTTCCGAGAGAGA 680
 Db 369 CACCTATCAGACACCCCTGTATGATTTCCAGTTGACAACTTCCAGCTTCCGAGAGAGA 428
 QY 681 TACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAAGTGGGCGAGCTCCGCTTC 740
 Db 429 TACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAAGTGGGCGAGCTCCGCTTC 488
 QY 741 ACTCTTGGCAGCCCTGGACATCACCACATCCGCTCAAGACCAATGGAGGGTGAATGG 800
 Db 489 ACTCTTGGCAGCCCTGGACATCACCACATCCGCTCAAGACCAATGGAGGGTGAATGG 548
 QY 801 GAAAGGGTCACTCAAGGAGACACCGCGGAGACATCTACCACAGACCTGGGCGAGACTT 860
 Db 549 GAAAGGGTCACTCAAGGAGACACCGCGGAGACATCTACCACAGACCTGGGCGAGACTT 608
 QY 861 TGTGAAGTTCCTGGATGCTATGTGTGAGCACAAGTTACAGTTCTGGGCGAGTGAAGTGA 920
 Db 609 TGTGAAGTTCCTGGATGCTATGTGTGAGCACAAGTTACAGTTCTGGGCGAGTGAAGTGA 668
 QY 921 AAATGAGCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTCCCTGGGCTTCAACCC 980
 Db 669 AAATGAGCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTCCCTGGGCTTCAACCC 728
 QY 981 TGAACATCAGCAGAGACTTCAATGCGCGTGACCTAGTCTACCTCGCCCAACAGTACTCA 1040
 Db 729 TGAACATCAGCAGAGACTTCAATGCGCGTGACCTAGTCTACCTCGCCCAACAGTACTCA 788
 QY 1041 CCACAAATGTCCGCTTACTCATGTCTGGATGACCAACAGCTTGTGCTGCCCGGCTGGGCAAA 1100
 Db 789 CCACAAATGTCCGCTTACTCATGTCTGGATGACCAACAGCTTGTGCTGCCCGGCTGGGCAAA 848

QY 1101 GGTGGTACTGACAGCCAGACCACTAAATATGTTTAT-GGCATGCTGTACATGGT 1159
 Db 849 GTGTGCTACTGACAGAACCCAAACAGCTTAATATGTTCTGGCATGCTGTAATGGG 908
 QY 1160 ACCTGGACTTT 1170
 Db 909 AACCTGGACTT 919
 RESULT 7
 BG744064 916 bp mRNA linear EST 15-MAY-2001
 LOCUS 602722846F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849307 5',
 DEFINITION mRNA sequence.
 ACCESSION BG744064
 VERSION BG744064.1 GI:14054717
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 916)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue procurement: Dr. Daniel McVicar, DBS/NCI
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1688 row: 1 column: 12
 High quality sequence stop: 856.
 Location/Qualifiers
 1..916
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4849307"
 /clone_lib="NIH_MGC_106"
 /tissue_type="natural killer cells, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: blood; Vector: pORF7; Site_1: xhoI; Site_2:
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using Zap-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."
 BASE COUNT 202 a 257 c 258 g 199 t
 ORIGIN
 Query Match 37.7%; Score 858.4; DB 12; Length 916;
 Best Local Similarity 99.2%; Pred. No. 1.6e-159;
 Matches 873; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 QY 1156 TGGTACTGAGCTTTCTGGCTCCAGCCAAAGCCACCTAGGGAGACACACCGCTGTC 1215
 Db 12 TGGTACTGAGCTTTCTGGCTCCAGCCAAAGCCACCTAGGGAGACACACCGCTGTC 71
 QY 1216 CCCAACACCATGCTCTTTTGGCTCAGAGCCCTGTTGGCTCCCAAGTTCGGGAGCAGAGT 1275
 Db 72 CCCAACACCATGCTCTTTTGGCTCAGAGCCCTGTTGGCTCCCAAGTTCGGGAGCAGAGT 131
 QY 1276 GTGGGCTAGGCTCTGGGATCGAGGATGACGTACAGCAGCAGCATCATCAGAACCTC 1335
 Db 132 GTGGGCTAGGCTCTGGGATCGAGGATGACGTACAGCAGCAGCATCATCAGAACCTC 191
 QY 1336 CTGTACATGTGGTGGCTGGACCGATGGAACCTTGCCCTGAACCCCGAAGAGGAGCC 1395
 Db 1336 CTGTACATGTGGTGGCTGGACCGATGGAACCTTGCCCTGAACCCCGAAGAGGAGCC 1395

Db 192 CTGTACCATGTGGTGGCTGGACCGAGCTTGGACCTTGCCTGAACCCCGAAGAGGAGCC 251
 QY 1396 AATGGGTGCGTAACTTTTGTGACAGTCCCATCATTTGTAGACATCACCACGAGACACCTTT 1455
 Db 252 AATGGGTGCGTAACTTTTGTGACAGTCCCATCATTTGTAGACATCACCACGAGACACCTTT 311
 QY 1456 TACAAACAGCCCATGTTCTTACCACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCC 1515
 Db 312 TACAAACAGCCCATGTTCTTACCACCTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCC 371
 QY 1516 CAGACAGTGGGGCTGTTGGCCACTTCAGCAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 1575
 Db 372 CAGACAGTGGGGCTGTTGGCCACTTCAGCAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 431
 QY 1576 CCGGATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1635
 Db 432 CCGGATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
 QY 1636 ATCAAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1695
 Db 492 ATCAAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 551
 QY 1696 TACCTGTGGCTGCCAGTGTGATGGAGCAGATCTCAAGGAGGAGGAGGAGGAGGAGGAGG 1755
 Db 552 TACCTGTGGCTGCCAGTGTGATGGAGCAGATCTCAAGGAGGAGGAGGAGGAGGAGGAGG 611
 QY 1756 CATTAAAGGAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1815
 Db 612 CATTAAAGGAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 671
 QY 1816 AGTGATGAGCTTACAGGACGATGAGCCAGGAGCAATGTTTGGGTGAGTCTGCTGCTGCTGCT 1875
 Db 672 AGTGATGAGCTTACAGGACGATGAGCCAGGAGCAATGTTTGGGTGAGTCTGCTGCTGCTGCT 731
 QY 1876 CTAGTGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935
 Db 732 CTAGTGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 791
 QY 1936 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1995
 Db 792 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 851
 QY 1996 GTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2034
 Db 852 GTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891
 RESULT 8
 BG744733/c
 LOCUS 602722846T1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849307 3',
 DEFINITION mRNA sequence.
 ACCESSION BG744733
 VERSION BG744733.1 GI:14055386
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 957)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue procurement: Dr. Daniel McVicar, DBS/NCI
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1688 row: 1 column: 12

Db	849	ACAGTTCTGGGCACTAACAGCTGAATAATGAGCCTTCTGCTGGCGCTGTGTGATGGATACACC	790
QY	957	CTTTCAGTGGCTGGGCTTACCCCTGAAACATACAGAGACTTCATTCGCCGTGACCTAGG	1016
Db	789	CTTTCAGTGGCT--GGCTTACCCCTGACATACAGAGACTTCATTCGCCGTGACCTAGG	731
QY	1017	TCCTACCCCTGGCAACAGTACACAGTACACCAATGTCCGCTTACATCATGTGTGATGACCAAG	1076
Db	730	TCCTACCCCTGGCAACAGACTACCAATGTCCGCTTACATCATGTGTGATGACCAAG	671
QY	1077	CTTGTCTGCTGCCCACTGGGCAAGGTGGTACTGACAGACCCAGAACGACGCTAAATATGT	1136
Db	670	CTTGTCTGCTGCCCACTGGGCAAGGTGGWACTKACAGACCCAGAACGACGCTAAATATGT	611
QY	1137	TCATGGCAATGTGTGTACATGTGTACCTGGACCTTTCTTGGCTCCAGCAAAAGCCACCTAGG	1196
Db	610	TCATGGCAATGTGTGTACATGTGTACCTGGACCTTTCTTGGCTCCAGCAAAAGCCACCTAGG	551
QY	1197	GGAGACACACCGGCTGTTCCCAACACATGCTTTTGCCCTCAGAGGCTGTGTGGGCTC	1256
Db	550	GAAGAACACACCGGCTGTTCCCAACACATGCTTTTGCCCTCAGAGGCTGTGTGGGCTC	491
QY	1257	CAAGTCTCTGGGACAGAGTGTGCGGCTAGGCTCTGGGATGAGGGATGACGATACAGCCA	1316
Db	490	CAAGTCTCTGGGACAGAGTGTGCGGCTARGCYCTGGGATCGAGGATGCAAAACACCA	431
QY	1317	CAGCATCATCAGAACCTCCTGTACCATGTGCTGGCTGGGACGACCTGGAACCTGCCCT	1376
Db	430	CAGCAYCATCAGAACCTCCTGTACCATGTGCTGGCTGGGACCACTGGAACCTGCCCT	371
QY	1377	GAACCCGAGGAGGACCACTGGGTCGCTTACCTTTGTCACAGTCCCATCATTTGTAGA	1436
Db	370	AAACCCGAGGAGGACCACTGGGTCGCTTACCTTTGTCACAGTCCCATCATTTGTAAA	311
QY	1437	CATCACCAAGGACAGCTTTTCAACAGACCCATGTTTCACCACTTGGGCACCTTCAGCAA	1496
Db	310	CATCACCAAGGACAGCTTTTCAACAGACCCATGTTTCACCACTTGGGCACCTTCAGCAA	251
QY	1497	GTTTCATTCCTGAGGGCTCCAGAGAGTGGGCTGGTTCGACAGTACAGAACGACCTGGA	1556
Db	250	GTTTCATTCCTGAGGGCTCCAGAGAGTGGGCTGGTTCGACATCAAAACACVACCTGGA	191
QY	1557	CGCAGTGGCACTGATCATCCGATGGCTCTGCTGTGTGTGGTGTGTGTAAACGGCTCTTC	1616
Db	190	CGCAGTGGCACTGATCATCCGATGGCTCTGCTGTGTGTGGTGTGTGTAAACGGCTCTTC	131
QY	1617	TAAAGATGTGGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTGTGAGCAATCTTCACC	1676
Db	130	TAAAGATGTGGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTGTGAGCAATCTTCACC	71
QY	1677	TGGTACTTCATTCACCTACCTGTGGCTCGCCAGTGGAGGACAGATCTCAAGGAG	1736
Db	70	TGGTACTTCATTCACCTACCTGTGGCTCGCCAGTGGGCGAGATCTCAAGGAG	11
QY	1737	GCATCTGG	1743
Db	10	GCMCTGG	4
RESULT 10			
B0680411			
LOCUS			
DEFINITION			
AGENCOURT_8209539 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6272642			
5', mRNA sequence.			
B0680411			
B0680411.1			
GI:21793090			
EST.			
SOURCE			
Homo sapiens			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 931)			
REFERENCE			
AUTHORS			
NIH-MGC http://mhc.nci.nih.gov/ .			

```
Db 541 AAGGGGTCACTCAAGGGACAGCCGGAGACATCTACCACAGACCTGGCCAGATACCTTT 600
Qy 862 GTGAAGTTCCTGATGCTATCTCAGACAGAGTTACAGTTCTGGGAGTGCAGAGCTGAA 921
Db 601 GTGAAGTTCCTGATGCTATCTCAGACAGAGTTACAGTTCTGGGAGTGCAGAGCTGAA 660
Qy 922 AATGAGCCTTCTGCTGGGCTGTGAGTGGATGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 981
Db 661 AATGAGCCTTCTGCTGGGCTGTGAGTGGATGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 720
Qy 982 GAACATCAGCAGAGATTCATGTCGCCGTGACCTAGTGTCTACCTGCGCAACAGTACTCAC 1041
Db 721 GAACATCAGCAGAGATTCATGTCGCCGTGACCTAGTGTCTACCTGCGCAACAGTACTCAC 779
Qy 1042 CACAATGTCGCCGTACTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1101
Db 780 CCAATGTCGCCGTACTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
Qy 1102 GTGGTACTGACAGACCCAGAGACCTAAATATGTTATGTCATGTCATGTCATGTCATGTCAT 1161
Db 840 G-GGGACTGACGACCCAGAGACCTAAATATGTCATGTCATGTCATGTCATGTCATGTCAT 898
Qy 1162 CTGGACTTCTCG 1174
Db 899 CCGGACTTTTCGG 911

RESULT 11
BQ681028 838 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8209132 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6262714
DEFINITION 5', mRNA sequence.
ACCESSION BQ681028
VERSION BQ681028.1 GI:21793707
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 838)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DPF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2425 row: h column: 11
High quality sequence stop: 716.
Location/Qualifiers
1..838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6262714"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 191 a 262 c 202 g 183 t
ORIGIN
```

```
Query Match 36.6%; Score 835; DB 14; Length 838;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 835; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 262 GTGGTGTGTCTGTGCATGCCACATACCTGTGACTCTCTTTGACCCCGGACCTTTCTTCC 321
Db 1 GTGGTGTGTCTGTGCATGCCACATACCTGTGACTCTCTTTGACCCCGGACCTTTCTTCC 60
Qy 322 CTTGGTACCTTCAGCCCTATGAGAGTACACGAGTGGGCGGAGTGGAGCTGAGTATG 381
Db 61 CTTGGTACCTTCAGCCCTATGAGAGTACACGAGTGGGCGGAGTGGAGCTGAGTATG 120
Qy 382 GGGCCCCATCCAGGCTAATACACGGGCGACAGGCTGTCTACTGACCTGCGACGAGAACAG 441
Db 121 GGGCCCCATCCAGGCTAATACACGGGCGACAGGCTGTCTACTGACCTGCGACGAGAACAG 180
Qy 442 AAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACAGATGCTGCTGCTCTCAACATC 501
Db 181 AAGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACAGATGCTGCTGCTCTCAACATC 240
Qy 502 CTTGCCCTGTCAACCCCTGCCAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 561
Db 241 CTTGCCCTGTCAACCCCTGCCAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 300
Qy 562 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGACCTAC 360
Qy 622 ACCTATCGACACACCCCTGTATTTCCAGTTCGCACAACTTCAGCTCCCGAGAGGAGAT 681
Db 361 ACCTATCGACACACCCCTGTATTTCCAGTTCGCACAACTTCAGCTCCCGAGAGGAGAT 420
Qy 682 ACCAAGCTCAAGATACACCCCTGATTCACGAGCCCTGAGTTGGCCAGCGCTCCCGTTTCA 741
Db 421 ACCAAGCTCAAGATACACCCCTGATTCACGAGCCCTGAGTTGGCCAGCGCTCCCGTTTCA 480
Qy 742 CTCCTTCCGACCCCTGGACATCACCCACTTTGGCTCAAGACCAATGAGCGGTGAATGGG 801
Db 481 CTCCTTCCGACCCCTGGACATCACCCACTTTGGCTCAAGACCAATGAGCGGTGAATGGG 540
Qy 802 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACCCACAGACCTGGGCGGACATCTTT 861
Db 541 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACCCACAGACCTGGGCGGACATCTTT 600
Qy 862 GTGAAGTTCCTGATGCTATGCTGAGCACAAGTTACAGTTCTGGGAGTGCAGAGCTGAA 921
Db 601 GTGAAGTTCCTGATGCTATGCTGAGCACAAGTTACAGTTCTGGGAGTGCAGAGCTGAA 660
Qy 922 AATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 981
Db 661 AATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 720
Qy 982 GAACATCAGCAGAGATTCATGTCGCCGTGACCTAGTGTCTACCTGCGCAACAGTACTCAC 1041
Db 721 GAACATCAGCAGAGATTCATGTCGCCGTGACCTAGTGTCTACCTGCGCAACAGTACTCAC 780
Qy 1042 CACAATGTCGCCGTACTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1096
Db 781 CACAATGTCGCCGTACTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 835
```

```
RESULT 12
BQ677036 916 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8209522 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6273665
DEFINITION 5', mRNA sequence.
ACCESSION BQ677036
VERSION BQ677036.1 GI:21789715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 916)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2451 row: g column: 02
High quality sequence stop: 535.
Location/Qualifiers
1..916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6272665"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/note="organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
into EcoRI/XhoI sites using the following 5' adaptor:
GCCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

Query Match 36.6%; Score 833.4; DB 14; Length 916;
Best Local Similarity 97.1%; Pred. No. 1.4e-154;
Matches 879; Conservative 0; Mismatches 23; Indels 3; Gaps 3;
Qy 262 GTGGTGTGTCTGCAATGCCACATATGTGACCTCTTTCACCCCGACCTTTCTCTGCC 321
Db 1 GTGGTGTGTCTGCAATGCCACATATGTGACCTCTTTCACCCCGACCTTTCTCTGCC 60
Qy 322 CTGGTACCTTCAGCGCTATGAGAGTACAGCGAGTGGCGGCGATGGAGTGTAGTATG 381
Db 61 CTGGTACCTTCAGCGCTATGAGAGTACAGCGAGTGGCGGCGATGGAGTGTAGTATG 120
Qy 382 GGGCCATCCAGCTAATCACACGGGCACAGCGCTGTCTACTGACCTTCAGCCGAGCAACAG 441
Db 121 GGGCCATCCAGCTAATCACACGGGCACAGCGCTGTCTACTGACCTTCAGCCGAGCAACAG 180
Qy 442 AGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACAGATGCTGCTCTCAACATC 501
Db 181 AGTTCAGAAAGTGAAGGATTTGGAGGGCCATGACAGATGCTGCTCTCAACATC 240
Qy 502 CTGGCTCTGACCCCTGCCAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 561
Db 241 CTGGCTCTGACCCCTGCCAAATTTGCTACTTAAATCGTACTTCTCTGAAGAAGGA 300
Qy 562 ATCGGATATAACATCATCCGGGTACCCATGCCAGCTGTGACTTTCATCCATCCGACCTAC 621
Db 301 ATCGGATATAACATCATCCGGGTACCCATGCCAGCTGTGACTTTCATCCATCCGACCTAC 360
Qy 622 ACCTATGACAGACCCCTGATGATTTCCAGTTCGACACTTACAGCTCCAGAGGAAGAT 681
Db 361 ACCTATGACAGACCCCTGATGATTTCCAGTTCGACACTTACAGCTCCAGAGGAAGAT 420
Qy 682 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGAGTGGCCAGCGTCCCGTTTCA 741
Db 421 ACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGAGTGGCCAGCGTCCCGTTTCA 480
Qy 742 CTCCTTGCAGCCCTTGACATCATCCACCTTGGCTCAAGACCAATGGAGCGGTGAATGGG 801
|||||

Db 481 CTCCTTGCAGCCCTTGAGCATCACCCACTTGGCTCAAGACCAATGAGGCGTGAATGGG 540
Qy 802 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACACACAGACCTGGCCAGATCTTT 861
Db 541 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACACACAGACCTGGCCAGATCTTT 600
Qy 862 GTGAAGTTCCTGGATGCTATGCTGAGCAGCAAGTTTACAGTTCTTGGCAGTGCAGCTGAA 921
Db 601 GTGAAGTTCCTGGATGCTATGCTGAGCAGCAAGTTTACAGTTCTTGGCAGTGCAGCTGAA 660
Qy 922 AATGAGCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCACCCCT 981
Db 661 AATGAGCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGGSCCTGGGCTTC-CCCT 719
Qy 982 GAACATCAGCAGACTTCATTTGCCGTGACCTAGTCTCCTACCTCCGACACAGTACTCAC 1041
Db 720 GAACATCAGCAGACTTCATTTGCCGTGACCTAGTCTCCTACCTCCGACACAGTACTCA- 778
Qy 1042 CACATCTCCGCTACTCATGCTGGATGACCAACGCTTGTCTGCTGCCCTGGGCAAG 1101
Db 779 CACATCTCCGCTACTCATGCTGGATGACCAACGCTTGTCTGCTGCCCTGGGCAAG 838
Qy 1102 GTGTACTACAGACCCAGCAGAGCAAGCAATTAATGTCTAT-GGCATTTGCTGTACATGGTA 1160
Db 839 GGGGAGCTGACAGCCAGACCAACCAAGCAATTAATGTCTATGGTCAATGGGCAATGGCTGGA 898
Qy 1161 CCTGG 1165
Db 899 CCTGG 903
RESULT 13
BM477383
LOCUS
DEFINITION AGENCOURT_6485207 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554173
5', mRNA sequence.
ACCESSION BM477383
VERSION BM477383.1 GI:18526425
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2272 row: e column: 22
High quality sequence stop: 674.
Location/Qualifiers
1..1017
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5554173"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 227 a 289 c 261 g 240 t
ORIGIN

Db 421 ACCAAGCTCAGATACCCCTGATTCACCGAGCCCTGCGAGTGGCCAGCGTCCCGTTTCA 480
 QY 742 CTCCTTCCAGCCCTGGGACATACCCCTGGCTCAAGACCAATGGAGCGGTGAATGGG 801
 Db 481 CTCCTTCCAGCCCTGGGACATACCCCTGGCTCAAGACCAATGGAGCGGTGAATGGG 540
 QY 802 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACCAACAGAGCTGGGCGAGATCTTT 861
 Db 541 AAGGGTCACTCAAGGACAGCCCGGAGACATCTACCAACAGAGCTGGGCGAGATCTTT 600
 QY 862 GTGAAGTTCCTGGATGCTATGCTGAGCACAGTTTACAGTTCTGGGCGAGTCAAGCTGAA 921
 Db 601 GTGAAGTTCCTGGATGCTATGCTGAGCACAGTTTACAGTTCTGGGCGAGTCAAGCTGAA 660
 QY 922 AATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGTCCCTGGGCTTCAACCCCT 981
 Db 661 AATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTCCAGTCCCTGGGCTTCAACCCCT 720
 QY 982 GAACATCAGGAGACTTCATTGCGGTGACCTAGTGTCTACCCCTGCGCAACAGTACTCAC 1041
 Db 721 GAACATCAGGAGACTTCATTGCGGTGACCTAGTGTCTACCCCTGCGCAACAGTACTCAC 780
 QY 1042 CACATGTCGCGCTACTGCTATGCTGATGACCAAGCTGTGCTGCTGCCCACTGGGCAAG 1101
 Db 781 CACATGTCGCGCTACTTCTGCTGATGACCAAGCTGTGCTGCTGCCCACTGGGCAAG 840
 QY 1102 GTGCTACTGACAC 1116
 Db 841 GGGTGGTACTGGAC 855

RESULT 15
 BQ069685
 LOCUS
 DEFINITION BQ069685.1 1008 bp mRNA linear EST 02-APR-2002
 AGENCOURT_7054386 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5921960
 5', mRNA sequence.
 BQ069685
 BQ069685.1 GI:19898731
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1008)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgep@rsfemail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCN2088 row: b column: 09
 High quality sequence stop: 696.
 Location/Qualifiers
 1..1008
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5921960"
 /clone.lib="NIH_MGC_47"
 /tissue_type="neuroblastoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOB7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dr priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library.
 BASE COUNT 228 a 293 c 250 g 236 t
 ORIGIN

Query Match 36.2%; Score 826; DB 14; Length 1008;
 Best Local Similarity 97.6%; Pred. No. 4e-153;
 Matches 891; Conservative 0; Mismatches 16; Indels 6; Gaps 5;

QY 1 ACCTAAGCGAGTACCTGTCATCCTCTGTTTGTAGTGGATCCTCTATCTCTTCAAGAC 60
 Db 29 AGCTAAGCGAGTACCTGTCATCCTCTGTTTGTAGTGGATCCTCTATCTCTTCAAGAC 88
 QY 61 TCTGGAACCCCTGTGGTCTTCTTCATCTAATGACCTGAGGGATGGAGTTTCAAGT 120
 Db 89 TCTGGAACCCCTGTGGTCTTCTTCATCTAATGACCTGAGGGATGGAGTTTCAAGT 148
 QY 121 CTTTCCAGAGAGGAATGTCCTCCAGCCCTTGAAGTAAAGCATCATGCTGCGACGCTC 180
 Db 149 CTTTCCAGAGAGGAATGTCCTCCAGCCCTTGAAGTAAAGCATCATGCTGCGACGCTC 208
 QY 181 ACAGATGCTTCTTACTTACGAGTGTGCTGGGATCAGTGTGCTGCTGCTGCTGCTGCT 240
 Db 209 ACAGATGCTTCTTACTTACGAGTGTGCTGGGATCAGTGTGCTGCTGCTGCTGCTGCT 268
 QY 241 AAMAGCTTCGGCTACAGCTCGGTGTGTCTGCAATGCCACATATCTGTGACTCCTTT 300
 Db 269 AAMAGCTTCGGCTACAGCTCGGTGTGTCTGCAATGCCACATATCTGTGACTCCTTT 328
 QY 301 GACCCCGGAGCTTCTCCTGCTTGGTACCTTTCAGCCGCTATGAGAGTACACGAGTGG 360
 Db 329 GACCCCGGAGCTTCTCCTGCTTGGTACCTTTCAGCCGCTATGAGAGTACACGAGTGG 388
 QY 361 CGAGGATGAGAGTGTAGTATGGGCGCATCCAGGCTAATACACGGGCACAGCCCTGCTA 420
 Db 389 CGAGGATGAGAGTGTAGTATGGGCGCATCCAGGCTAATACACGGGCACAGCCCTGCTA 448
 QY 421 CTGACCTTCGAGCGAGACAGAGTTCAGAGAGTGAAGGATTTGGAGGGGCGCATGACA 480
 Db 449 CTGACCTTCGAGCGAGACAGAGTTCAGAGAGTGAAGGATTTGGAGGGGCGCATGACA 508
 QY 481 GATGCTGCTGCTCTACATCCTTGCCTGTGTCACCCCTGCGGCAAAATTTGCTACTTAA 540
 Db 509 GATGCTGCTGCTCTACATCCTTGCCTGTGTCACCCCTGCGGCAAAATTTGCTACTTAA 568
 QY 541 TCGTACTTCTCTGAAGAAGATTCGATATAACATCATTCGGGTACCATGCGGCGCTGT 600
 Db 569 TCGTACTTCTCTGAAGAAGATTCGATATAACATCATTCGGGTACCATGCGGCGCTGT 628
 QY 601 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCAAC 660
 Db 629 GACTTCTCCATCCGACCTACACCTATGCAGACACCCCTGATGATTTCCAGTTGCAAC 688
 QY 661 TTCAGCTTCCAGAGGAAGATACCAAGTCAAGATACCCCTGATTCACGAGCCCTGCAAG 720
 Db 689 TTCAGCTTCCAGAGGAAGATACCAAGTCAAGATACCCCTGATTCACGAGCCCTGCAAG 748
 QY 721 TTGCGCCAGCGTCCGCTTTCCTCTTCCAGCCCTGAGACATCACCCCTGGCTGCAAG 780
 Db 749 TTGCGCCAGCGTCCGCTTTCCTCTTCCAGCCCTGAGACATCACCCCTGGCTGCAAG 808
 QY 781 ACCAATGAGCGCT-GAATGGAGAGGCTCACTCAAGGGA-CAGCCCGGAGACATCTACC 838
 Db 809 AACAATGGAGCGGGAATGGGAAGGGTCACTCAGGGACGACCCGCGAGACATCTACC 868
 QY 839 ACCAG--ACCTGGGCGAGATCTTTGTGAAGTTCT--GGATGCTATGCTG-AGCACAG 894
 Db 869 ACCAGGACCTGGGCGGACGATCTTTGTGAAGTTCTTGGGATGCTGCTGTAAGCACAG 928
 QY 895 TTACAGTTCTGGG 907
 Db 929 TTACAGTTCTCGG 941

Mon Feb 24 10:51:34 2003

us-10-024-197-24.rst

Page 14

Search completed: February 22, 2003, 13:08:25
Job time : 2957 secs

...

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 10:33:37 ; Search time 102 Seconds
(without alignments)
6852.121 Million cell updates/sec

Title: US-10-024-197-24

Perfect score: 2279

Sequence: 1 agctaagcaggtacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/ptodata/1/ina/5C.COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6A.COMB.seq.*
5: /cgn2.6/ptodata/1/ina/6B.COMB.seq.*
6: /cgn2.6/ptodata/1/ina/6C.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1601.6	70.3	1642	2	US-08-713-928B-3
2	667	29.3	7620	1	US-07-767-135-1
3	667	29.3	7620	1	US-07-841-652-1
4	89.4	3.9	358	1	US-07-841-652-2
5	77	3.4	89	1	US-07-767-135-19
6	74	3.2	75	1	US-07-767-135-2
7	49.6	2.2	7218	1	US-08-232-463-14
8	42.2	1.9	289	4	US-09-007-005-17
9	42.2	1.9	289	4	US-09-244-796-17
10	41	1.8	41	1	US-08-317-432A-44
11	40.8	1.8	7218	1	US-08-232-463-14
12	39.4	1.7	41	1	US-08-317-432A-46
13	37.4	1.6	41	1	US-08-317-432A-50
14	36.6	1.6	379	3	US-09-026-343-21
15	36.6	1.6	379	3	US-09-362-871-21
16	36.6	1.6	379	3	US-09-362-871-21
17	36	1.6	53526	3	US-08-658-136-2
18	36	1.6	53526	3	US-08-658-136-1
19	35.2	1.5	11219	3	US-07-642-734C-1
20	35.2	1.5	11219	3	US-08-439-009A-1
21	34.8	1.5	1424	4	US-09-326-186B-226
22	34.8	1.5	1428	5	PCT-US94-09642-1
23	34.8	1.5	1957	1	US-08-295-060-3
24	34.2	1.5	198	1	US-08-330-108-16
25	34.2	1.5	198	5	PCT-US92-10087-16
26	34	1.5	4143	4	US-08-826-134-6
27	34	1.5	5350	4	US-08-826-134-3

28	33.8	1.5	1952	1	US-08-333-359-1	Sequence 1, Appl
29	33.8	1.5	1952	1	US-08-463-694-1	Sequence 1, Appl
30	33.8	1.5	1952	1	US-08-694-501-1	Sequence 1, Appl
31	33.8	1.5	2004	1	US-08-471-033-18	Sequence 18, Appl
32	33.8	1.5	2004	2	US-08-471-033-18	Sequence 18, Appl
33	33.8	1.5	2004	2	US-08-463-483A-18	Sequence 18, Appl
34	33.8	1.5	2004	2	US-08-463-483A-18	Sequence 18, Appl
35	33.8	1.5	2004	2	US-08-470-566B-18	Sequence 18, Appl
36	33.8	1.5	2004	2	US-08-469-334-18	Sequence 18, Appl
37	33.8	1.5	2004	3	US-09-300-529-35	Sequence 35, Appl
38	33.8	1.5	2576	1	US-08-471-033-35	Sequence 35, Appl
39	33.8	1.5	2576	2	US-08-471-044-35	Sequence 35, Appl
40	33.8	1.5	2576	2	US-08-463-483A-35	Sequence 35, Appl
41	33.8	1.5	2576	2	US-08-471-046A-35	Sequence 35, Appl
42	33.8	1.5	2576	2	US-08-470-566B-35	Sequence 35, Appl
43	33.8	1.5	2576	2	US-08-469-334-35	Sequence 35, Appl
44	33.8	1.5	2576	3	US-09-300-529-35	Sequence 35, Appl
45	33.8	1.5	2655	1	US-08-471-033-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-713-928B-3
; Sequence 3, Application US/08713928B
; Patent No. 5929304
; GENERAL INFORMATION:
; APPLICANT: RADIN, DAVID N.
; APPLICANT: CRAMER, CAROLE L.
; APPLICANT: OISHI, KAREN K.
; APPLICANT: WEISSENBOERN, DEBORAH L.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/713,928B
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/003,737
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7956-0011-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1642 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: CDNA
; MOLECULE TYPE: CDNA
US-08-713-928B-3
Query Match 70.3%; Score 1601.6; DB 2; Length 1642;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1604; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 106 ATGGAGTTTCAAGTCCCTCCAGAGAGGAGTCTCCCAAGCCTTTGAGTAGGTAGCATC 165
DB 1 ATGGAGTTTCAAGTCCCTCCAGAGAGGAGTCTCCCAAGCCTTTGAGTAGGTAGCATC 60

QY 166 ATGGCTGGCAGGCTCACAGGATTCCTTACCTTACAGGAGTGTCTGGGCGCATCAGGTGCC 225
DB 61 ATGGCTGGCAGGCTCACAGGATTCCTTACCTTACAGGAGTGTCTGGGCGCATCAGGTGCC 120

QY 226 CGCCCTTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGAAATGCCACA 285
DB 121 CGCCCTTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGAAATGCCACA 180

QY 286 TACTGTGACTCTTGTACCCCGGACCTTCTCTGCTTGTGCTTGTACCTTCAGCGCTATGAG 345
DB 181 TACTGTGACTCTTGTACCCCGGACCTTCTCTGCTTGTGCTTGTGCTTGTGCTTGTGAG 240

QY 346 AGTACAGGAGTGGCGGAGGATGGAGCTGAGTATGGGCGCCATCCAGGCTAATCACAG 405
DB 241 AGTACAGGAGTGGCGGAGGATGGGCTGAGTATGGGCGCCATCCAGGCTAATCACAG 300

QY 406 GGCACAGGCTGTACTGACCTGACCGAGGAGTCCAGAGAGTCCAGAAAGTGAAGGATTT 465
DB 301 GGCACAGGCTGTACTGACCTGACCGAGGAGTCCAGAGAGTCCAGAAAGTGAAGGATTT 360

QY 466 GGAGGGCCATGACAGATGCTGCTCTCAACATCTTGGCCCTGTACCCCTGCCCAA 525
DB 361 GGAGGGCCATGACAGATGCTGCTCTCAACATCTTGGCCCTGTACCCCTGCCCAA 420

QY 526 AATTGCTACTTAATTCGATCTCTGAAAGAGGATTCGGAATATCAATCATCCGGGTA 585
DB 421 AATTGCTACTTAATTCGATCTCTGAAAGAGGATTCGGAATATCAATCATCCGGGTA 480

QY 586 CCATGCGGAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACCCCTGATGAT 645
DB 481 CCATGCGGAGCTGTGACTTCTCCATCCGACCTACACCTATGACAGACCCCTGATGAT 540

QY 646 TTCAGCTTGCACAACTTACGCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGAT 705
DB 541 TTCAGCTTGCACAACTTACGCTCCAGAGGAGATACCAAGCTCAAGATACCCCTGAT 600

QY 706 CACCGAGCCTGGAGTTGGCCAGAGTCCGCTTCTACTCTTCCAGCCCTGGACATCA 765
DB 601 CACCGAGCCTGGAGTTGGCCAGAGTCCGCTTCTACTCTTCCAGCCCTGGACATCA 660

QY 766 CCACCTTGGCTCAAGACCAATGAGGCGTGAATGGAGGGGTCACTCAAGGAGACGCC 825
DB 661 CCACCTTGGCTCAAGACCAATGAGGCGTGAATGGAGGGGTCACTCAAGGAGACGCC 720

QY 826 GGAGACATCTACACAGACCTGGGCGGATCTTTGTGAAATTCCTGGATGCCATGCT 885
DB 721 GGAGACATCTACACAGACCTGGGCGGATCTTTGTGAAATTCCTGGATGCCATGCT 780

QY 886 GAGCACAAGTTACAGTTCTGGGAGTACAGCTGAAATGAGCCTTCTCTGGCTGTG 945
DB 781 GAGCACAAGTTACAGTTCTGGGAGTACAGCTGAAATGAGCCTTCTCTGGCTGTG 840

QY 946 AGTGGATACCCCTTCCAGTGGCTTCCAGCTTACCCCTGAACATCAGGAGACTTCATG 1005
DB 841 AGTGGATACCCCTTCCAGTGGCTTCCAGCTTACCCCTGAACATCAGGAGACTTCATG 900

QY 1006 CGTGACCTAGGCTTACCTCGGCAACAGTACTACCAACATGTCCGCTTACTATGCTG 1065
DB 901 CGTGACCTAGGCTTACCTCGGCAACAGTACTACCAACATGTCCGCTTACTATGCTG 960

QY 1066 GATGACCAAGCTTGTGCTGCCCGGAGTGGGAGGAGTGGTACTGACAGACCCAGACA 1125
DB 961 GATGACCAAGCTTGTGCTGCCCGGAGTGGGAGGAGTGGTACTGACAGACCCAGACA 1020

QY 1126 GCTAAATATGTTTACGCAATGCTGTACATGCTGACTGACTGCTGCTCCAGCCAAA 1185
DB 1021 GCTAAATATGTTTACGCAATGCTGTACATGCTGACTGACTGCTGCTCCAGCCAAA 1080

RESULT 2

US-07-767-135-1
; Sequence 1, Application US/07767135
; Patent No. 5234811
; GENERAL INFORMATION:
; APPLICANT: Beutler, Ernest
; APPLICANT: Sorge, Joseph A
; TITLE OF INVENTION: An Assay for a New Gaucher Disease
; TITLE OF INVENTION: Mutation
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 3366 No. 5234811th Torrey Pines Court, Suite 240
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07767,135
; FILING DATE: 19910927
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bingham, Douglas A
; REGISTRATION NUMBER: 32,457
; REFERENCE/DOCKET NUMBER: SCR0422P
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7620 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: exon
LOCATION: 356..611
OTHER INFORMATION: /product= "Exon 1"
FEATURE:
NAME/KEY: intron
LOCATION: 612..979
OTHER INFORMATION: /function= "Intron 1"
FEATURE:
NAME/KEY: exon
LOCATION: 980..1067
OTHER INFORMATION: /product= "Exon 2"
FEATURE:
NAME/KEY: intron
LOCATION: 1068..1619
OTHER INFORMATION: /function= "Intron 2"
FEATURE:
NAME/KEY: exon
LOCATION: 1620..1811
OTHER INFORMATION: /product= "Exon 3"
FEATURE:
NAME/KEY: intron
LOCATION: 1812..1934
OTHER INFORMATION: /function= "Intron 3"
FEATURE:
NAME/KEY: exon
LOCATION: 1935..2081
OTHER INFORMATION: /product= "Exon 4"
FEATURE:
NAME/KEY: intron
LOCATION: 2082..3046
OTHER INFORMATION: /function= "Intron 4"
FEATURE:
NAME/KEY: exon
LOCATION: 3047..3180
OTHER INFORMATION: /product= "Exon 5"
FEATURE:
NAME/KEY: intron
LOCATION: 3181..3390
OTHER INFORMATION: /function= "Intron 5"
FEATURE:
NAME/KEY: exon
LOCATION: 3391..3563
OTHER INFORMATION: /product= "Exon 6"
FEATURE:
NAME/KEY: intron
LOCATION: 3564..4116
OTHER INFORMATION: /function= "Intron 6"
FEATURE:
NAME/KEY: exon
LOCATION: 4117..4354
OTHER INFORMATION: /product= "Exon 7"
FEATURE:
NAME/KEY: intron
LOCATION: 4355..5227
OTHER INFORMATION: /function= "Intron 7"
FEATURE:
NAME/KEY: exon
LOCATION: 5228..5452
OTHER INFORMATION: /product= "Exon 8"
FEATURE:
NAME/KEY: intron

LOCATION: 5453..5852
OTHER INFORMATION: /function= "Intron 8"
FEATURE:
NAME/KEY: exon
LOCATION: 5853..6016
OTHER INFORMATION: /product= "Exon 9"
FEATURE:
NAME/KEY: intron
LOCATION: 6017..6385
OTHER INFORMATION: /function= "Intron 9"
FEATURE:
NAME/KEY: exon
LOCATION: 6386..6502
OTHER INFORMATION: /product= "Exon 10"
FEATURE:
NAME/KEY: intron
LOCATION: 6503..6596
OTHER INFORMATION: /function= "Intron 10"
FEATURE:
NAME/KEY: exon
LOCATION: 6597..7245
OTHER INFORMATION: /function= "Exon 11"
US-07-787-135-1
Query Match 29.3%; Score 667; DB 1; Length 7620;
Best Local Similarity 89.1%; Pred. No. 2e-179;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;
QY 1489 TTCAGCAAGTTTCATTCCTGAGGGCTCCCGAGAGTGGGGCTGGTGGCCAGTCAGAGAAAC 1548
|||||
Db 6381 TTCAGCAAGTTTCATTCCTGAGGGCTCCCGAGAGTGGGGCTGGTGGCCAGTCAGAGAAAC 6440
QY 1549 GACCTGGACGACGTGGCACTGATGCCATCCGATGGCTCTGCTGTGTGGTGGTCTGCTAAAC 1608
|||||
Db 6441 GACCTGGACGACGTGGCACTGATGCCATCCGATGGCTCTGCTGTGTGGTGGTCTGCTAAAC 6500
QY 1609 C-----
Db 6501 CGGTGAGGCAATGTTGAGTCTGGGAAGTGGGTGAAGACAGCGCTTGGGGCCCTTGGCA 6560
QY 1610 -----GCTCCTTAAGGATGTGCTCTTAC 1634
Db 6561 GGATCACACTCTCAGCTTCTCTCCTGCTCCTAGCTCCTTAAGGATGTGCTCTTAC 6620
QY 1635 CATCAAGGATCCTGCTGTGGGCTTCTCGAGACAATCTACCTGGCTACTCCATTACAC 1694
|||||
Db 6621 CATCAAGGATCCTGCTGTGGGCTTCTCGAGACAATCTACCTGGCTACTCCATTACAC 6680
QY 1695 CTACCTGTGGCGTCCGACGTGATGGAGCAGATCTCAAGGAGGCACTGGGGCTCAGCCTGG 1754
|||||
Db 6681 CTACCTGTGGCGTCCGACGTGATGGAGCAGATCTCAAGGAGGCACTGGGGCTCAGCCTGG 6740
QY 1755 GCATTAAGGGACAGAGTCAGCTCACAGCTGTCTGTGACTAAAGAGGGCAGCAGAGGGC 1814
Db 6741 GCATTAAGGGACAGAGTCAGCTCACAGCTGTCTGTGACTAAAGAGGGCAGCAGAGGGC 6800
QY 1815 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGGTTGGGTGACTCACTTTCCCC 1874
Db 6801 CAGTGTGAGCTTACAGCGACGTAAGCCAGGGGCAATGGTTGGGTGACTCACTTTCCCC 6860
QY 1875 TCTAGGTGGTCCAGGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCGAGTGTCCCC 1934
Db 6861 TCTAGGTGGTCCAGGGGCTGGAGGCCCTAGAAAAGATCAGTAAGCCCGAGTGTCCCC 6920
QY 1935 CCAGCCCCATGCTTATGTGAACATGCCCTGTGTGCTGTGCTTTGGAACCTGGGCTG 1994
Db 6921 CCAGCCCCATGCTTATGTGAACATGCCCTGTGTGCTGTGCTTTGGAACCTGGGCTG 6980
QY 1995 GGTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACACAAAGATCAGGCTGAGGCTAAG 2054
Db 6981 GGTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACACAAAGATCAGGCTGAGGCTAAG 7040
QY 2055 GAAAAGAGAGACTAGGAAAGCTGGGCCCCAAACTGGAGACTGTTTGTCTTCTCTGGAGA 2114

Db 7041 GAAAGAGAGAGACTAGGAAAGCTGGGCCCAAACTGGAGAGCTGTTGCTTCTCTGGAGA 7100
QY 2115 TCAGAACTGGCCCGTGGAGCAGCAGTGTGAGATCAGGCGGGAAGCCTTAAAGCAGCA 2174
Db 7101 TCAGAACTGGCCCGTGGAGCAGCAGTGTGAGATCAGGCGGGAAGCCTTAAAGCAGCA 7160
QY 2175 GGGGTGTGCCAGCAGCAGCAGATGATTCCTATGCGCAGCAGCAGGAAAATGGCAGCTC 2234
Db 7161 GCGGTGTGCCAGCAGCAGCAGATGATTCCTATGCGCAGCAGCAGGAAAATGGCAGCTC 7220
QY 2235 TTAAGGAGAAAATCTTTGAGCCCA 2259
Db 7221 TTAAGGAGAAAATCTTTGAGCCCA 7245

RESULT 3

US-07-841-652-1
; Sequence 1, Application US/07841652
; Patent No. 5266459
; GENERAL INFORMATION:
; APPLICANT: Beutler, Ernest
; TITLE OF INVENTION: GAUCHER'S DISEASE: DETECTION OF A NEW
; TITLE OF INVENTION: MUTATION IN INTRON 2 OF THE GLUCOCEREBROSIDASE GENE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5266459th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,652
; FILING DATE: 19920224
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bingham, Douglas A
; REGISTRATION NUMBER: 32,457
; REFERENCE/DOCKET NUMBER: SCR0670P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7620 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: exon
; LOCATION: 356..611
; OTHER INFORMATION: /product= "Exon 1"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 612..979
; OTHER INFORMATION: /function= "Intron 1"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 980..1067
; OTHER INFORMATION: /product= "Exon 2"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1068..1619
; OTHER INFORMATION: /function= "Intron 2"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1620..1811
; OTHER INFORMATION: /product= "Exon 3"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1812..1934
; OTHER INFORMATION: /function= "Intron 3"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1935..2081
; OTHER INFORMATION: /product= "Exon 4"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2082..3046
; OTHER INFORMATION: /function= "Intron 4"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3047..3180
; OTHER INFORMATION: /product= "Exon 5"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3181..3390
; OTHER INFORMATION: /function= "Intron 5"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3391..3563
; OTHER INFORMATION: /product= "Exon 6"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3564..4116
; OTHER INFORMATION: /function= "Intron 6"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4117..4354
; OTHER INFORMATION: /product= "Exon 7"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 4355..5227
; OTHER INFORMATION: /function= "Intron 7"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5228..5452
; OTHER INFORMATION: /product= "Exon 8"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 5453..5852
; OTHER INFORMATION: /function= "Intron 8"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5853..6016
; OTHER INFORMATION: /product= "Exon 9"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 6017..6385
; OTHER INFORMATION: /function= "Intron 9"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6386..6502
; OTHER INFORMATION: /product= "Exon 10"
; FEATURE:
; NAME/KEY: intron
; LOCATION: 6503..6596
; OTHER INFORMATION: /function= "Intron 10"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6597..7245
; OTHER INFORMATION: /product= "Exon 11"
US-07-841-652-1

Query Match 29.3%; Score 567; DB 1; Length 7620;
Best Local Similarity 89.1%; Pred. No. 2e-179;
Matches 771; Conservative 0; Mismatches 0; Indels 94; Gaps 1;

QY 1489 TTCAGCAAGTTCATCTCTGAGGGCTCCAGAGAGTGGGGTGTGCTGCCAGTCAGAAGAAC 1548
Db 6381 TTCAGCAAGTTCATCTCTGAGGGCTCCAGAGAGTGGGGTGTGCTGCCAGTCAGAAGAAC 6440
QY 1549 GACCTGGACGACGTGGCACTGATGCATCCCGATGGCTCTCTCTTTGTGTGCTGCTAAAC 1608
Db 6441 GACCTGGACGACGTGGCACTGATGCATCCCGATGGCTCTCTCTTTGTGTGCTGCTAAAC 6500
QY 1609 C----- 1609
Db 6501 CGGTGAGGGCAATGTTGAGTCTGGGAAGTGGGCTGAAGACAGCGTTGGGGCCCTTGCCA 6560
QY 1610 -----GCTCCTCTAAGGATGTGCCCTTTAC 1634
Db 6561 GGATCACACTCTCAGCTTCTCCTCCTGCTCCTCCTAGCTCCTTAAGCATGTGCCCTTTAC 6620
QY 1635 CATCAAGATCTCTGCTGTGGGCTTCTCGAGACAATCTCACCTGGCTACTCATTACACAC 1694
Db 6621 CATCAAGATCTCTGCTGTGGGCTTCTCGAGACAATCTCACCTGGCTACTCATTACACAC 6680
QY 1695 CTACCTGTGGCTGCGCATGATGAGCAGATCTCAAGCAGGCACTGGGCTGACCTCTGG 1754
Db 6681 CTACCTGTGGCTGCGCATGATGAGCAGATCTCAAGCAGGCACTGGGCTGACCTCTGG 6740
QY 1755 GCATTAAAGGACAGATCAGCTACAGCGCTCTGTGACTAAAGAGGCGCAGCAGGGC 1814
Db 6741 GCATTAAAGGACAGATCAGCTACAGCGCTCTGTGACTAAAGAGGCGCAGCAGGGC 6800
QY 1815 CAGTGTGAGCTTACAGCAGCTAAGCCAGCGGCAATGTTGGTGTGACTCACTTTCCCC 1874
Db 6801 CAGTGTGAGCTTACAGCAGCTAAGCCAGCGGCAATGTTGGTGTGACTCACTTTCCCC 6860
QY 1875 TCTAGTGTGTCAGGCGGCTGGAGGCCCTAGAAAAAGATCAGTAAAGCCCAAGTGTCCCC 1934
Db 6861 TCTAGTGTGTCAGGCGGCTGGAGGCCCTAGAAAAAGATCAGTAAAGCCCAAGTGTCCCC 6920
QY 1935 CCAGCCCCATGCTTATGTGAACATCGCTGTGCTGTGCTTGTGTTGGAACCTGGGGCTG 1994
Db 6921 CCAGCCCCATGCTTATGTGAACATCGCTGTGCTGTGCTTGTGTTGGAACCTGGGGCTG 6980
QY 1995 GGTCCAGGCTTAGGTTAGCTCAGCTGTCCGTACAAACACAGATCAGGCGCTGAGGGTAAAG 2054
Db 6981 GGTCCAGGCTTAGGTTAGCTCAGCTGTCCGTACAAACACAGATCAGGCGCTGAGGGTAAAG 7040
QY 2055 GAAAAAGAGACTAGAAAGCTGGGCCCAAACTGGAGACTCTTGTCTTCTTCCTGGAGA 2114
Db 7041 GAAAAAGAGACTAGAAAGCTGGGCCCAAACTGGAGACTCTTGTCTTCTTCCTGGAGA 7100
QY 2115 TGCAGACTGGGCCCTGGAGCAGAGTGCAGATCAGGGCGGAAGCCTTAAAGCAGCA 2174
Db 7101 TGCAGACTGGGCCCTGGAGCAGAGTGCAGATCAGGGCGGAAGCCTTAAAGCAGCA 7160
QY 2175 GCGGTGTGCCAGGACCCAGATGTTCTTATGTCACAGCAGGCAAAATGCGAGCTC 2234
Db 7161 GCGGTGTGCCAGGACCCAGATGTTCTTATGTCACAGCAGGCAAAATGCGAGCTC 7220
QY 2235 TTAAGGAGAAATGTTTGAAGCCCA 2259
Db 7221 TTAAGGAGAAATGTTTGAAGCCCA 7245

RESULT 4
US-07-841-652-2
; Sequence 2, Application US/07841652
; Patent No. 5266459
; GENERAL INFORMATION:
; APPLICANT: Beutler, Ernest
; TITLE OF INVENTION: GAUCHER'S DISEASE: DETECTION OF A NEW
; TITLE OF INVENTION: MUTATION IN INTRON 2 OF THE GLUCOCEREBROSIDASE GENE

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/07/767.135
FILING DATE: 19910927
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bingham, Douglas A
REGISTRATION NUMBER: 32,457
REFERENCE/DOCKET NUMBER: SCR0422P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 89 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-767-135-19

Query Match 3.4%; Score 77; DB 1; Length 89;
Best Local Similarity 98.9%; Pred. No. 1.2e-12;
Matches 88; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 133 GAATGTCCTCCAGCCCTTTGAGTAGGTAAGCATCATGGCTGCAGCCTCACAGGATT-GCT 191
|||||
Db 1 GAATGTCCTCCAGCCCTTTGAGTAGGTAAGCATCATGGCTGCAGCCTCACAGGATTGGCT 60
QY 192 TCTACTTCAGGCAGTGCTGGGCATCAG 220
|||||
Db 61 TCTACTTCAGGCAGTGCTGGGCATCAG 89

RESULT 6
US-07-767-135-2
Sequence 2, Application US/07767135
Patent No. 5234811
GENERAL INFORMATION:
APPLICANT: Beutler, Ernest
APPLICANT: Sorge, Joseph A
TITLE OF INVENTION: An Assay for a New Gaucher Disease
TITLE OF INVENTION: Mutation
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 3366 No. 5234811th Torrey Pines Court, Suite 240
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/767.135
FILING DATE: 19910927
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bingham, Douglas A
REGISTRATION NUMBER: 32,457

REFERENCE/DOCKET NUMBER: SCR0422P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 57
OTHER INFORMATION: /note= "N is either G in a normal
OTHER INFORMATION: glucocerebrosidase gene or N is GG in a mutant
OTHER INFORMATION: glucocerebrosidase gene"
US-07-767-135-2

Query Match 3.2%; Score 74; DB 1; Length 75;
Best Local Similarity 98.7%; Pred. No. 7.7e-12;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 133 GAATGTCCTCCAGCCCTTTGAGTAGGTAAGCATCATGGCTGCAGCCTCACAGGATTGCTT 192
|||||
Db 1 GAATGTCCTCCAGCCCTTTGAGTAGGTAAGCATCATGGCTGCAGCCTCACAGGATTNCTT 60
QY 193 CTACTTCAGGCAGTG 207
|||||
Db 61 CTACTTCAGGCAGTG 75

RESULT 7
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 2.2%; Score 49.6; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.00071;
Matches 4; Conservative 164; Mismatches 88; Indels 0; Gaps 0;
QY 2024 GTACAAACAAGATCAGGCTGAGGCTTAAGGAAAGAGAGACTAGGAAAGCTGGGCC 2083
DB 1438 GTACRR 1379
QY 2084 AAAACTGGAGACTGTTGCTTCTCGGAGATGAGAACTGGGCCGCTGGAGCAGCATG 2143
DB 1378 RRR 1319
QY 2144 TCACCATCAGCGGAGGCTTAAGCAGCAGCGGCTGCCAGCAGCCACCATCATC 2203
DB 1318 RRR 1259
QY 2204 CTATGGCAGCAGCAGAAATGCGACTCTTAAAGGAGAAATGTTTTCAGCCCAAAA 2263
DB 1258 RRR 1199
QY 2264 AAAAAAAGAAAAA 2279
DB 1198 RRRRRRRRRRRRR 1183

RESULT 8
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.9%; Score 42.2; DB 4; Length 289;
Best Local Similarity 12.5%; Pred. No. 0.017;
Matches 33; Conservative 95; Mismatches 137; Indels 0; Gaps 0;
QY 2015 TCACCTGCTCAACACAGATCAGGCTGAGGCTTAAGGAAAGAGAGACTAGGAAA 2074
DB 18 URARCRURARURARCRARARURARCRARURARCRURARCRURARCRURCRGR 77

QY 2075 GCTGGGCCCAAACTGGAGACTGTTGCTTCTCGGAGATGACAGAACTGGGCCCTGGA 2134
DB 78 NRSNR 137
QY 2135 GCAGCAGTGTACGATCAGCGGGAAGCCCTTAAAGCAGCAGCGGCTGCCAGGCACC 2194
DB 138 NRSNR 197
QY 2195 AGATGATTCCTATGGCAGCAGCAGGAAATGCGACTCTTAAAGCAGGAAATGTTTGA 2254
DB 198 NRSNR 257
QY 2255 GCCCAAAAAAAGAAAAA 2279
DB 258 CHUAAAAAAGAAAAA 282

RESULT 9
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.9%; Score 42.2; DB 4; Length 289;
Best Local Similarity 12.5%; Pred. No. 0.017;
Matches 33; Conservative 95; Mismatches 137; Indels 0; Gaps 0;

QY 2015 TCACCTGCTCAACACAGATCAGGCTGAGGCTTAAGGAAAGAGAGACTAGGAAA 2074
DB 18 URARCRURARURARCRARARURARCRARURARCRURARCRURARCRURCRGR 77
QY 2075 GCTGGGCCCAAACTGGAGACTGTTGCTTCTCGGAGATGACAGAACTGGGCCCTGGA 2134
DB 78 NRSNR 137
QY 2135 GCAGCAGTGTACGATCAGCGGGAAGCCCTTAAAGCAGCAGCGGCTGCCAGGCACC 2194
DB 138 NRSNR 197
QY 2195 AGATGATTCCTATGGCAGCAGCAGGAAATGCGACTCTTAAAGCAGGAAATGTTTGA 2254
DB 198 NRSNR 257
QY 2255 GCCCAAAAAAAGAAAAA 2279
DB 258 CHUAAAAAAGAAAAA 282

;; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
;; STREET: 2940 Birchtree lane
;; CITY: Silver Spring
;; STATE: Maryland
;; COUNTRY: United States of America
;; ZIP: 20906
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
;; COMPUTER: Twinhead* Slimnote-890TX
;; OPERATING SYSTEM: MS DOS version 6.2,
;; SOFTWARE: Word for Windows version 3.11
;; SOFTWARE: Word for Windows version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/317,432A
;; FILING DATE: 4-Oct-94
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/919,872
;; FILING DATE: 27-Jul-92
;; APPLICATION NUMBER: 08/084,505
;; FILING DATE: 1-Jul-93
;; NAME: Friedman, Mark M.
;; REGISTRATION NUMBER: 33,883
;; REFERENCE/DOCKET NUMBER: 128/7
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 972-3-562553
;; TELEFAX: 972-3-562554
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 46:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 41
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-317-432A-46
Query Match 1.7%; Score 39.4; DB 1; Length 41;
Best Local Similarity 97.6%; Pred. No. 0.039; 1; Indels 0; Gaps 0;
Matches 40; Conservative 0; Mismatches 1;
QY 1533 TGCCAGTCAGAGAACGACCTGGCGAGTCGACTGATGC 1573
DB 1 TGCCAGTCAGAGAACGACCTGGCGAGTCGACTGATGC 41
RESULT 13
US-08-317-432A-48/c
; Sequence 48, Application US/08317432A
; Patent No. 5710028
; GENERAL INFORMATION:
; APPLICANT: Nurit Eyal and Nir Navot
; TITLE OF INVENTION: A method of quick screening and
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; SOFTWARE: Word for Windows version 3.11
; SOFTWARE: Word for Windows version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,432A
; FILING DATE: 4-Oct-94
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 435
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41
; TYPE: nucleic acid

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/919,872
;; FILING DATE: 27-Jul-92
;; APPLICATION NUMBER: 08/084,505
;; FILING DATE: 1-Jul-93
;; NAME: Friedman, Mark M.
;; REGISTRATION NUMBER: 33,883
;; REFERENCE/DOCKET NUMBER: 128/7
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 972-3-562553
;; TELEFAX: 972-3-562554
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 48:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 41
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-317-432A-48
Query Match 1.7%; Score 39; DB 1; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.05; 0; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 0;
QY 1535 CCAGTCAGAGAACGACCTGGCGAGTCGACTGATGC 1573
DB 39 CCAGTCAGAGAACGACCTGGCGAGTCGACTGATGC 1
RESULT 14
US-08-317-432A-50/c
; Sequence 50, Application US/08317432A
; Patent No. 5710028
; GENERAL INFORMATION:
; APPLICANT: Nurit Eyal and Nir Navot
; TITLE OF INVENTION: A method of quick screening and
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; SOFTWARE: Word for Windows version 3.11
; SOFTWARE: Word for Windows version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,432A
; FILING DATE: 4-Oct-94
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/919,872
; FILING DATE: 27-Jul-92
; APPLICATION NUMBER: 08/084,505
; FILING DATE: 1-Jul-93
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 128/7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
; TELEX:
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41
; TYPE: nucleic acid

100

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 12:19:34 ; Search time 176 Seconds

(without alignments)
7272.592 Million cell updates/sec

Title: US-10-024-197-24

Perfect score: 2279

Sequence: 1 agctagcgagctacctgca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US03_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1785.6	78.4	1792	10	US-09-782-378A-16
2	1543	67.7	1551	10	US-09-753-126-2
3	667	29.3	75270	9	US-09-790-852-1
4	639.4	28.1	5769	10	US-09-964-824A-284
5	225.4	9.9	362	10	US-09-960-352-5649
c 6	130	5.7	135	9	US-10-046-935-1845
c 7	130	5.7	135	9	US-09-878-178-1845
c 8	39.6	1.7	375	10	US-09-960-352-7227
9	38.4	1.7	477	10	US-09-867-701-3268
10	38.4	1.7	2845	10	US-09-833-381-1497
11	38.4	1.7	3612	10	US-09-935-291A-10
12	38.4	1.7	3619	10	US-09-833-381-1495
13	38.4	1.7	3919	10	US-09-935-291A-8
14	36.6	1.6	379	12	US-10-028-780-21
c 15	36.4	1.6	293	10	US-09-880-107-1230
16	36	1.6	564	10	US-09-833-381-1498
17	36	1.6	53522	9	US-09-904-968A-1
18	35.8	1.6	964	10	US-09-764-864-300
c 19	35.4	1.6	314	10	US-09-867-701-10324

c 20	35.4	1.6	476	12	US-10-001-843-87	Sequence 87, Appl
21	35.2	1.5	509	10	US-09-867-701-5694	Sequence 5694, Ap
22	35	1.5	295	10	US-09-867-701-8588	Sequence 8588, Ap
23	34.8	1.5	1424	9	US-09-954-531-366	Sequence 366, App
24	34.8	1.5	1424	9	US-09-441-411-21	Sequence 21, Appl
25	34.8	1.5	1424	9	US-09-962-436-556	Sequence 556, App
c 26	34.8	1.5	155074	9	US-10-026-188-6	Sequence 6, Appl
27	34.6	1.5	559	9	US-09-883-152-27	Sequence 27, Appl
28	34.6	1.5	623	9	US-09-883-152-29	Sequence 29, Appl
29	34.6	1.5	687	9	US-09-774-639-106	Sequence 106, App
c 30	34.6	1.5	1665	10	US-09-821-167-6	Sequence 6, Appl
31	34.6	1.5	4308	9	US-10-042-431-79	Sequence 19, Appl
32	34.6	1.5	4308	9	US-09-756-130B-449	Sequence 449, App
c 33	34.6	1.5	11188	10	US-09-821-167-1	Sequence 1, Appl
34	34.2	1.5	15175	12	US-10-028-056-1	Sequence 1, Appl
c 35	34.2	1.5	15843	10	US-09-764-869-2396	Sequence 2396, Ap
c 36	34.2	1.5	16774	10	US-09-764-869-2395	Sequence 2395, Ap
c 37	34.2	1.5	16774	10	US-09-764-869-2398	Sequence 2398, Ap
c 38	34	1.5	319	10	US-09-867-701-2350	Sequence 2350, Ap
c 39	34	1.5	1798	10	US-09-925-299-230	Sequence 230, App
c 40	34	1.5	3945	10	US-09-917-800A-1566	Sequence 1566, Ap
c 41	34	1.5	53542	10	US-09-801-574-61	Sequence 61, Appl
c 42	33.8	1.5	1267	12	US-10-001-843-45	Sequence 45, Appl
c 43	33.6	1.5	398	10	US-09-960-352-12846	Sequence 12846, A
c 44	33.6	1.5	815	10	US-09-894-998-1	Sequence 1, Appl
c 45	33.6	1.5	1401	9	US-09-738-626-3465	Sequence 3465, Ap

ALIGNMENTS

RESULT 1
US-09-782-378A-16
; Sequence 16: Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-16

Query Match	78.4%	Score	1785.6	DB	10	Length	1792
Best Local Similarity	99.8%	Pred. No.	0				
Matches	1788	Conservative	0	Mismatches	4	Indels	0
Gaps	0						
QY 121	CCCTCCAGAGAGAGATGCTCCCAAGCCTTTGAGTAGGTAAGCATCATGCTGGCAGCCTC	180					
Db 1	CCCTCCAGAGAGAGATGCTCCCAAGCCTTTGAGTAGGTAAGCATCATGCTGGCAGCCTC	60					
QY 181	ACAGGATTGCTTCTACTTTCAGGAGTGTCTGTGGGATCATGCTGGCAGCCTC	240					
Db 61	ACAGGATTGCTTCTACTTTCAGGAGTGTCTGTGGGATCATGCTGGCAGCCTC	120					
QY 241	AAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTGCAATGCCACATACGTGACTCTTTT	300					
Db 121	AAAAGCTTCGGCTACAGCTCGGTGGTGTCTGTGCAATGCCACATACGTGACTCTTTT	180					
QY 301	GACCCCGACCTTCCCTGCGCTTGGTACCTTCAGCGGCTATGAGAGTACAGGAGTGGG	360					
Db 181	GACCCCGACCTTCCCTGCGCTTGGTACCTTCAGCGGCTATGAGAGTACAGGAGTGGG	240					

Qy	361	CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCGCTGCTA	420
Db	241	CGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCGCTGCTA	300
Qy	421	CTGACCCCTGCAGCCAGAACAGAAAGTTCAGAAAGTGAAGGATTTTGAGGGGCCATGACA	480
Db	301	CTGACCCCTGCAGCCAGAACAGAAAGTTCAGAAAGTGAAGGATTTTGAGGGGCCATGACA	360
Qy	481	GATGCTGCTGCTCTMACATTCCTTGCCCTGTACCCCTGCCAAAATTTTGCTACTTAAA	540
Db	361	GATGCTGCTGCTCTCACATTCCTTGCCCTGTACCCCTGCCAAAATTTTGCTACTTAAA	420
Qy	541	TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT	600
Db	421	TCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGT	480
Qy	601	GACTTCTCCATCCGCACTTACACCTATCAGACACCCCTGATGATTTCCAGTTGCAACAAC	660
Db	481	GACTTCTCCATCCGCACTTACACCTATCAGACACCCCTGATGATTTCCAGTTGCAACAAC	540
Qy	661	TTGAGCCTCCAGAGAGAAATACCAAGCTCAAGATACCCCTGATCATCCAGAGCCCTGCAG	720
Db	541	TTGAGCCTCCAGAGAGAAATACCAAGCTCAAGATACCCCTGATCATCCAGAGCCCTGCAG	600
Qy	721	TTGGCCCAAGCGTCCCGTTTCACTCCTTGCCAGCCCTGGACATCACCCATTGGCTCAAG	780
Db	601	TTGGCCCAAGCGTCCCGTTTCACTCCTTGCCAGCCCTGGACATCACCCATTGGCTCAAG	660
Qy	781	ACCAATGGAGCGGTGAATGGGAAGGGTCTACTCAAGGACAGCCCGGAGACATCTACCAC	840
Db	661	ACCAATGGAGCGGTGAATGGGAAGGGTCTACTCAAGGACAGCCCGGAGACATCTACCAC	720
Qy	841	CAGACCTGGCCAGATACCTTGTGAAGTTCTCTGGATGCTATGCTGAGCAACAAGTTACAG	900
Db	721	CAGACCTGGCCAGATACCTTGTGAAGTTCTCTGGATGCTATGCTGAGCAACAAGTTACAG	780
Qy	901	TTCTGGGCAGTGCAGCTGAAATGAGCCTTCTGCTGGGCTGTTTGAGTGGATACCCCTTC	960
Db	781	TTCTGGGCAGTGCAGCTGAAATGAGCCTTCTGCTGGGCTGTTTGAGTGGATACCCCTTC	840
Qy	961	CAGTGCCTGGGCTTACCCCTTGAACATCAGCCAGACTTCAATGGCCCTGACCTAGGTCCT	1020
Db	841	CAGTGCCTGGGCTTACCCCTTGAACATCAGCCAGACTTCAATGGCCCTGACCTAGGTCCT	900
Qy	1021	ACCTCGCCAAACAGTACTCACCAATGTCCGCTTACTCTGCTGGATGACCAACGCTTG	1080
Db	901	ACCTCGCCAAACAGTACTCACCAATGTCCGCTTACTCTGCTGGATGACCAACGCTTG	960
Qy	1081	CTGCTGCCCACTGGGAAAGTGTGTGACAGACACCAGAGCAGCTAAATGTGTCAT	1140
Db	961	CTGCTGCCCACTGGGAAAGTGTGTGACAGACACCAGAGCAGCTAAATGTGTCAT	1020
Qy	1141	GGCAATGCTGTACATTTGTAATGTAACCTTTCGGCTTCAGCCAAAGCCACCTAGGGGAG	1200
Db	1021	GGCAATGCTGTACATTTGTAATGTAACCTTTCGGCTTCAGCCAAAGCCACCTAGGGGAG	1080
Qy	1201	ACACACGGCCTGTTTCCCAACACCATGCTCTTTTGCTTCAGAGGCTGTGTGGGCTCAAG	1260
Db	1081	ACACACGGCCTGTTTCCCAACACCATGCTCTTTTGCTTCAGAGGCTGTGTGGGCTCAAG	1140
Qy	1261	TTCTGGGAGCAGAGTGTGGGGCTAGGCTCCTGGGATCAGGGATGCAGTACAGCCACAGC	1320
Db	1141	TTCTGGGAGCAGAGTGTGGGGCTAGGCTCCTGGGATCAGGGATGCAGTACAGCCACAGC	1200
Qy	1321	ATCATCAGAACTTCCTGTATACATGTGTGGCTGGACCGACTGGAACCTTGGCCTGAAC	1380
Db	1201	ATCATCAGAACTTCCTGTATACATGTGTGGCTGGACCGACTGGAACCTTGGCCTGAAC	1260
Qy	1381	CCCGAAGGAGACCCAAATTTGGGTGCGTAACATTTGTGCGAGTGCCTCATTTGTAGACATC	1440
Db	1261	CCCGAAGGAGACCCAAATTTGGGTGCGTAACATTTGTGCGAGTGCCTCATTTGTAGACATC	1320

2. WILLIS

```

US-09-753-126-2
; Sequence 2, Application US/09753126
; Patent No. US20020127219A1
; GENERAL INFORMATION:
; APPLICANT: OKKELS, JENS SIGURD
; APPLICANT: JENSEN, ANNE DAM
; APPLICANT: HALKIER, TORBEN
; APPLICANT: JENSEN, RIKKE BOLDING
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME
; TITLE OF INVENTION: ACTIVATORS
; FILE REFERENCE: 31-000600US
; CURRENT APPLICATION NUMBER: US/09753,126
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PA 1999 01891
; PRIOR FILING DATE: 1999-12-30
; PRIOR APPLICATION NUMBER: 60/174,652
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PA 200 00865
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/210,984
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 60/211,124
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: PA 2000 01027
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/217,497
; PRIOR FILING DATE: 2000-07-11
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-753-126-2

```

Query Match	67.7%;	Score 1543;	DB 10;	Length 1551;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1546; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

QY 1635 CATCAAGGATCTGCTGTGGGCTTCTTGAGACAAATCTCACTGGCTACTCCATTCACAC 1694
DB 38308 CATCAAGGATCTGCTGTGGGCTTCTTGAGACAAATCTCACTGGCTACTCCATTCACAC 38367
QY 1695 CTACCTGTGGGCTCGCCAGTGTATGAGCAGATPACTCAAGAGGACACTGGGCTCAGCCCTGG 1754
DB 38368 CTACCTGTGGGCTCGCCAGTGTATGAGCAGATPACTCAAGAGGACACTGGGCTCAGCCCTGG 38427
QY 1755 GCATTAAAGGGACAGTCACTCAGCTCAGCTGTCTGTACTTAAAGAGGACACAGGCGC 1814
DB 38428 GCATTAAAGGGACAGTCACTCAGCTCAGCTGTCTGTACTTAAAGAGGACACAGGCGC 38487
QY 1815 CAGTGTGAGCTTACAGCAGCTCAAGCCAGGCGCAATGTTTGGGTGACTCAGCTTTTCCCC 1874
DB 38488 CAGTGTGAGCTTACAGCAGCTCAAGCCAGGCGCAATGTTTGGGTGACTCAGCTTTTCCCC 38547
QY 1875 TCTAGTGTGTCAGGGGCTGAGGCGCTTAGAGCCCTTAGAAAAAGATCAGTAAGCCCGAGTGTCCCC 1934
DB 38548 TCTAGTGTGTCAGGGGCTGAGGCGCTTAGAGCCCTTAGAAAAAGATCAGTAAGCCCGAGTGTCCCC 38607
QY 1935 CCAGCCCATGCTTATGTGAACATGCGCTGTGCTGCTTGGAACTGGGCGCTG 1994
DB 38608 CCAGCCCATGCTTATGTGAACATGCGCTGTGCTGCTTGGAACTGGGCGCTG 38667
QY 1995 GTTCAGGCGCTAGGCTGAGCTCAGCTGTCCGTACAAACACAGATCAGGCGCTGAGGGTAAG 2054
DB 38668 GTTCAGGCGCTAGGCTGAGCTCAGCTGTCCGTACAAACACAGATCAGGCGCTGAGGGTAAG 38727
QY 2055 GAAAGACAGACTAGGAAGCTGGGCGCCCAAACTGGAGACTGTTTCTTCTTGGAGA 2114
DB 38728 GAAAGACAGACTAGGAAGCTGGGCGCCCAAACTGGAGACTGTTTCTTCTTGGAGA 38787
QY 2115 TGCAGAACTGGGCGCTGAGCAGCAGTGTGAGCTCAGGCGGGAAGCTTAAAGCAGCA 2174
DB 38788 TGCAGAACTGGGCGCTGAGCAGCAGTGTGAGCTCAGGCGGGAAGCTTAAAGCAGCA 38847
QY 2175 GCGGGTGTCCCGAGCCAGCCAGATGATTCCTATGCGACACAGCCAGGAAAAATGGCAGTCTC 2234
DB 38848 GCGGGTGTCCCGAGCCAGCCAGATGATTCCTATGCGACACAGCCAGGAAAAATGGCAGTCTC 38907
QY 2235 TTAAGGAGAAAAATGTTTGGAGCCCA 2259
DB 38908 TTAAGGAGAAAAATGTTTGGAGCCCA 38932

RESULT 4

US-09-964-824A-284
; Sequence 284, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrihan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 284
; LENGTH: 5769
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(5769)
; OTHER INFORMATION: n=a,t,g or c

US-09-964-824A-284

Query Match 28.1%; Score 639.4; DB 10; Length 5769;
Best Local Similarity 87.6%; Pred. No. 9.9e-170;
Matches 758; Conservative 0; Mismatches 12; Indels 95; Gaps 2;

QY 1489 TTCAAGCAAGTTTCAATTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTTGCCAGTCAAGAAAC 1548
DB 4792 TTCAAGCAAGTTTCAATTCCTGAGGGCTCCCAAGAGAGTGGGGCTGGTTGCCAGTCAAGAAAC 4851
QY 1549 GACCTGCACGACGAGTGGCAGTGCATCCGATGGCTCTCTGTGTGGTGTGGTCAAAAC 1608
DB 4852 GACCCGACGACGAGTGGCAGTGCATCCGATGGCTCTCTGTGTGGTGTGGTCAAAAC 4911
QY 1609 C----- 1609
DB 4912 CGGTGAGGGCAATGGTGAGGTCTGGGAAGTGGGCTGAAGACAGCGTTGGGGGCGCTTGGCA 4971
QY 1610 -----GCTCTCTAAAGATGTGCTCTTAC 1634
DB 4972 GATCAGCACTCTCAGCTTCTCCTCCTCCTCCTAGCTCCTCTAAAGGATGTGCTCTTAC 5031
QY 1635 CATCAAGGATCTGCTGTGGGCTTCTTGAGACAAATCTCAGCTGGCTTACTCCATTCACAC 1694
DB 5032 CATCAAGGATCTGCTGTGGGCTTCTTGAGACAAATCTCAGCTGGCTTACTCCATTCACAC 5091
QY 1695 CTACCTGTGGCGTCCGAGTGTAGGAGCAGATCTCAAGAGGACACTGGGCTCAGGCTGG 1754
DB 5092 CTACCTGTGGCGTCCGAGTGTAGGAGCAGATCTCAAGAGGACACTGGGCTCAGGCTGG 5151
QY 1755 GCATTAAAGGACAGATCAGCTCAGCTCAGCTGTGCTGTGACTTAAAGAGGACAGCAGGCGC 1814
DB 5152 GCATTAAAGGACAGATCAGCTCAGCTCAGCTGTGCTGTGACTTAAAGAGGACAGCAGGCGC 5211
QY 1815 CAGTGTGAGCTTACAGCGACGTAAGCCAGGCGCAATGTTTGGGTGACTCAGCTTTTCCCC 1874
DB 5212 CAGTGTGAGCTTACAGCGACGTAAGCCAGGCGCAATGTTTGGGTGACTCAGCTTTTCCCC 5271
QY 1875 TCTAGTGTGTCAGGCGCTGGAGGCGCTTAGAAAAAGATCAGTAAGCCCGAGTGTCCCC 1934
DB 5272 TCTAGTGTGTCAGGCGCTGGAGGCGCTTAGAAAAAGATCAGTAAGCCCGAGTGTCCCC 5331
QY 1935 CCAGCCCATGCTTATGTGAACATGCGCTGTGCTGTGCTTGTGAAACTGGGCGCTG 1994
DB 5332 CCAGCCCATGCTTATGTGAACATGCGCTGTGCTGTGCTTGTGAAACT--GGCGCTG 5390
QY 1995 GGTCCAGGCGCTAGGCTGAGCTCAGCTGTCCGTACAAACACAGATCAGGGCTGAGGGTAAG 2054
DB 5391 GGTCCAGGCGCTAGGCTGAGCTCAGCTGTCCGTACAAACACAGATCAGGGCTGAGGGTAAG 5450
QY 2055 GAAAGAAAGAGACTAGGAAAGCTGGGCGCCAAACTGGAGACTGTTTCTTCTTCTGAGA 2114
DB 5451 GAAAGAAAGAGACTAGGAAAGCTGGGCGCCAAACTGGAGACTGTTTCTTCTTCTGAGA 5510
QY 2115 TGCAGAACTGGGCGCTGGAGCAGCAGTGTGCTCAGCATCAGGCGGGAAGCTTAAAGCAGCA 2174
DB 5511 TNNNNNCTGGGCGCTGGAGCAGCAGTGTGAGCATCAGGGCGGAAGCTTAAAGCAGCA 5570
QY 2175 GCGGGTGTGCCAGGCGCCAGATGATTCCTATGGCAGCCAGGAAAAATGGCAGCTC 2234
DB 5571 GCGGGTGTGCCAGGCGCCAGATGATTCCTATGGCAGCCAGGAAAAATGGCAGCTC 5630
QY 2235 TTAAGGAGAAAAATGTTTGGAGCCCA 2259
DB 5631 TTAAGGAGAAAAATGTTTGGAGCCCA 5655

RESULT 5

US-09-960-352-5649
; Sequence 5649, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 5649
LENGTH: 362
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 24-LIB34-044-Q1-E1-F7
US-09-960-352-5649

Query Match 9.9%; Score 225.4; DB 10; Length 362;
Best Local Similarity 78.1%; Pred. No. 1.1e-53;
Matches 271; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 712 GCCCTGAGTTGGCCAGCGTCCGTTTTCACCTCTTGCAGCCCTGGACATCAACCCACT 771
DB 16 GCAGCCACCAAGCCCAACCGCTGACTCCTCTTGCAGTCCCTGGACATCAACCCACT 75
QY 772 TGGCTCAAGACCAATGAGCGGTGAATGGAGGGTCTCACTCAAGGGACAGCCCGGAGAC 831
DB 76 TGGCTCAAGACTAATGGGCTGTGAATGGAGGGGACACTCAAGGGTCAAGCGAGAT 135
QY 832 ATCTACCAACAGACCTGGGCCAGATCTTTGTGAAGTTCTTGGATGCTATGCTGAGCAC 891
DB 136 CTCTACCAACAGACCTGGGCCAGATCTTTGTCAAGTTCTTGGATGCTATGCTGAGCAC 195
QY 892 AAGTTACAGTTCTGGCCAGTGCAGCTGAAATGAGCTTCTGCTGGGCTGTGAGTGA 951
DB 196 AAGTTACAGTTCTGGCCAGTGCAGCTGAAATGAGCTTCTGCTGGGCTGTGAGTGA 255
QY 952 TACCCCTTCAGTGCCTGGGCTTCAACCTGAACATCAGCAGACTTCATTGCCCGGTGAC 1011
DB 256 TACCCCTTCAGTGCCTGGGCTTCAACCTGAACATCAGCAGACTTCATTGCCCGGTGAC 315
QY 1012 CTAGTCTCTACCTCGCCCAACAGTACTCACCACATGTCCGCCCTACT 1058
DB 316 CTGGGGTCTCAACCGTCCGCAAGAGACACGACCGCTACGTCGCACT 362

RESULT 6
US-10-046-935-1845/c
Sequence 1845, Application US/10046935
Patent No. US20020156011A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yuqiu
APPLICANT: Harlocker, Susan L.
APPLICANT: Secrist, Heather
APPLICANT: Wang, Aijun
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.527C1
CURRENT APPLICATION NUMBER: US/10/046,935
CURRENT FILING DATE: 2002-01-15
NUMBER OF SEQ ID NOS: 2239
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1845
LENGTH: 135
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46, 52, 53
OTHER INFORMATION: n = A,T,C or G
US-10-046-935-1845

Query Match 5.7%; Score 130; DB 9; Length 135;

Best Local Similarity 97.7%; Pred. No. 4.9e-27;
Matches 130; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1880 GTGTGCCAGGGGCTGGAGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCCCCAGC 1939
DB 135 GTGTGCCAGGGGCTGGAGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCCCCAGC 76
QY 1940 CCCCATGCTTATGTGAACATCGCTGTGCTTCTTTGGAACTGGGCTGGGTCC 1999
DB 75 CCCCATGCTTATGTGAACATCGCTGTGCTTCTTTGGAACTGGGCTGGGTCC 16
QY 2000 AGGCCTAGGGTGA 2012
DB 15 AGGCCTAGGGTGA 3

RESULT 7
US-09-878-178-1845/c
Sequence 1845, Application US/09878178
Patent No. US2002017752A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yuqiu
APPLICANT: Harlocker, Susan L.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.527
CURRENT APPLICATION NUMBER: US/09/878,178
CURRENT FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 2237
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1845
LENGTH: 135
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)-(135)
OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1845

Query Match 5.7%; Score 130; DB 9; Length 135;
Best Local Similarity 97.7%; Pred. No. 4.9e-27;
Matches 130; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1880 GTGTGCCAGGGGCTGGAGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCCCCAGC 1939
DB 135 GTGTGCCAGGGGCTGGAGCCCTAGAAAAAGATCAGTAAGCCCACTGTCCCCCAGC 76
QY 1940 CCCCATGCTTATGTGAACATCGCTGTGCTTCTTTGGAACTGGGCTGGGTCC 1999
DB 75 CCCCATGCTTATGTGAACATCGCTGTGCTTCTTTGGAACTGGGCTGGGTCC 16
QY 2000 AGGCCTAGGGTGA 2012
DB 15 AGGCCTAGGGTGA 3

RESULT 8
US-09-960-352-7227/c
Sequence 7227, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112

```
; SEQ ID NO 7227
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 31-LTB3058-041-Q1-K1-H3
US-09-960-352-7227

Query Match
Best Local Similarity 1.7%; Score 39.6; DB 10; Length 376;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 985 CATCAGCAGACTTCAATGCGCGTGCACCTAGTCTACCTCGCCACACAGTACTCACCAC 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 CACCTGGTAGTGAGCTGGCGCGTCCATAGCGCCACACAGCCGCGGCGGCGGACAG 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1045 AATGTCGCGCTACTCATGCTGGATGACCAACGCTTGTCTGCTGCCCACTGGGCAAAAGTG 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 CGTGCGCGCGACACACAGCCAGAACCCAGCGCGCGGTACCAAGGCGGCTGCGGG 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1105 GTACTGACAGCCAGACGACCTAAATATGTTCAATGGCATTGCTGTACATTTGTTACCTG 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 GTCGCGGACACCATGAGGAGTCTCGGAAGTCCACGCTCTTCAGGTGCATGCCCTCGG 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1165 GACTTTCGGCTCCAGCAAAAGCCACCTAGGG 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 GGCCTCCAGGTAGTGTCCAGGCCCTCGTTGGCG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-867-701-3268
; Sequence 3268, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867.701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3268
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-3268

Query Match
Best Local Similarity 1.7%; Score 38.4; DB 10; Length 477;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1457 ACAAACAGCCCATGTTCTACCACTTGCCACTTTCAGCAAGTTTCATCTCTGAGGCTCCC 1516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 ACCCCAGCAGCTGCTCCGCTTCATCCCTCATCTGTCAGAGTGTTCGCCCTGTGGCTCCCA 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1517 AGAGAGTGGGCTGGTTGCCAGTACAGAAAGACGACCTGGACGCACTGGCAGTGCATC 1576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 AGCAGAAGGAGTTTGTATCATCACCAGCCCTGAAGGAGCTGGGTACGTACCTCATGTGTG 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1577 CGATGGCTCTGCTGTTGGTGTGCTGTAAACCCGCTCTCTAAGGATGTGCTCTTACCA 1636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 GGGATGGCACCACGACGTGGCGGCCCTGAAGAGTGTGACGTGGGTGTGGCTCTTGG 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1637 TCAAGGATCCTG 1648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 CCAATGCCCTG 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-833-381-1497
; Sequence 1497, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1497
; LENGTH: 2846
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1497

Query Match
Best Local Similarity 1.7%; Score 38.4; DB 10; Length 2846;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1457 ACAAACAGCCCATGTTCTACCACTTGCCACTTTCAGCAAGTTTCATCTCTGAGGCTCCC 1516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1445 ACCCCAGCAGCTGCTCCGCTTCATCCCTCATCTGTCAGAGTGTTCGCCCTGTGGCTCCCA 1504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1517 AGAGAGTGGGCTGGTTGCCAGTACAGAAAGACGACCTGGACGAGTGGCAGTGCATC 1576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1505 AGCAGAAGGAGTTTGTATCATCACCAGCCCTGAAGGAGCTGGGTACGTACCTCATGTGTG 1564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1577 CGATGGCTCTGCTGTTGGTGTGCTGTAAACCCGCTCTCTAAGGATGTGCTCTTACCA 1636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1565 GGGATGGCACCACGAGCTGGCGGCCCTGAAGCATGCTGACGTGGGTGTGGCTCTTGG 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1637 TCAAGGATCCTG 1648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1625 CCAATGCCCTG 1636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-935-291A-10
; Sequence 10, Application US/09935291A
; Patent No. US20020107373A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chun, Mlyoung
; TITLE OF INVENTION: 49937, 49931, AND 49933, NOVEL HUMAN TRANSPORTER
; FILE REFERENCE: MNI-184
; CURRENT APPLICATION NUMBER: US/09/935,291A
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/226,504
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: 60/250,932
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3612
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-935-291A-10

Query Match
Best Local Similarity 1.7%; Score 38.4; DB 10; Length 3612;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1457 ACAAACAGCCCATGTTCTACCACTTGCCACTTTCAGCAAGTTTCATCTCTGAGGCTCCC 1516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2468 ACCCCAGCAGCTGCTCCGCTTCATCCCTCATCTGTCAGAGTGTTCGCCCTGTGGCTCCCA 2527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1517 AGAGAGTGGGCTGGTTGCCAGTACAGAAAGACGACCTGGACGAGTGGCAGTGCATC 1576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2528 AGCAGAAGGAGTTTGTATCATCACCAGCCCTGAAGGAGCTGGGTACGTACCTCATGTGTG 2587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1577 CCGATGCTCTGCTGTGTGGTCGTAAACCGCTCTCTAAGGATGGCTCTTACCA 1636
Db 2588 GGGATGGCCACCAACGAGCTGGGGCCCTGAAGCATGCTGAGTGGGTGTGGCGCTCTGG 2647

QY 1637 TCAAGGATCTG 1648
Db 2648 CCAATGCCCTG 2659

RESULT 12
US-09-833-381-1495
; Sequence 1495, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR FILING DATE: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1495
; LENGTH: 3639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1495

Query Match 1.7%; Score 38.4; DB 10; Length 3639;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1457 ACAAAACAGCCCATGTTCTACCACTTGCCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCC 1516
Db 2399 ACCCCAGCAGCTGCTCCGCTCATCCCCCATGTGCAGGTGTTCCGCCGTGGCTCCCA 2458

QY 1517 AGAGAGTGGGCTGTGGTCCAGTCAGAACAGACCTGGAGCGAGTGGCACTGATGCATC 1576
Db 2459 AGCAGAAGGAGTTTGCATCACCACCTGGAAGAGCTGGCGCTACGTGACCTCATGTGTG 2518

QY 1577 CCGATGCTCTGCTGTGTGGTCGTAAACCGCTCTCTAAGGATGGCTCTTACCA 1636
Db 2519 GGGATGGCCACCAACGAGCTGGGGCCCTGAAGCATGCTGAGTGGGTGGCGCTCTTGG 2578

QY 1637 TCAAGGATCTG 1648
Db 2579 CCAATGCCCTG 2590

RESULT 13
US-09-935-291A-8
; Sequence 8, Application US/09935291A
; Patent No. US20020107373A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chun, Miyoung
; TITLE OF INVENTION: 49937, 49931, AND 49933, NOVEL HUMAN TRANSPORTER
; FILE REFERENCE: MNI-184
; CURRENT APPLICATION NUMBER: US/09/935,291A
; CURRENT FILING DATE: 2001-08-21
; PRIOR FILING DATE: 60/226,504
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: 60/250,932
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 3919
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: (68)..(3679)
US-09-935-291A-8

Query Match 1.7%; Score 38.4; DB 10; Length 3919;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1457 ACAAAACAGCCCATGTTCTACCACTTGCCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCC 1516
Db 2535 ACCCCAGCAGCTGCTCCGCTCATCCCCCATGTGCAGGTGTTCCGCCGTGGCTCCCA 2594

QY 1517 AGAGAGTGGGCTGTGGTCCAGTCAGAACAGACCTGGAGCGAGTGGCACTGATGCATC 1576
Db 2595 AGCAGAAGGAGTTTGCATCACCAGCTGAAGGAGCTGGCGCTACGTGACCTCATGTGTG 2654

QY 1577 CCGATGCTCTGCTGTGTGGTCGTAAACCGCTCTCTAAGGATGGCTCTTACCA 1636
Db 2655 GGGATGGCCACCAACGAGCTGGGGCCCTGAAGCATGCTGAGTGGGTGGCGCTCTTGG 2714

QY 1637 TCAAGGATCTG 1648
Db 2715 CCAATGCCCTG 2726

RESULT 14
US-10-028-780-21
; Sequence 21, Application US/10028780
; Patent No. US20020132329A1
; GENERAL INFORMATION:
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: SHILATIFARD, ALI
; APPLICANT: CONAWAY, JOAN W.
; APPLICANT: CONAWAY, RONALD C.
; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSER: STERNE:
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/028,780
; FILING DATE: 28-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,343
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 60/038,447
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1488.0880001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 372-2540
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-028-780-21

Query Match      1.6%; Score 36.6; DB 12; Length 379;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 90; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1504 CCTGAGGCTCCACAGAGTGGGCTGGTGGCCAGTCAGAGAGACGACCTGGAGCCAGTG 1563
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13  CCTGAAGGCTACAGAAGACCGGAGCTACTTGTCTAGACTCCAGAAAGATGGTGTCAATCAA 72

QY 1564 GCACGTGATGCATCCGATGGCTCTGCTCTTGTGGTCTAAACCGCTCCTCTAAGGAT 1623
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 73  AAAGACAGAAGACTCCTCGGAGCAATCTGCAACAGGTAGCCATCTGAATTCCTAAGGAC 132

QY 1624 GTGCCCTTTACCATCAAGATCCTGCTGTGGGCTTCCTGGAGACAATCTCACCTGGCTAC 1683
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 CTTCTATATACCTTAAGGATTATGTTTTTAAAGAGCTTCAAGAGAGACTGGCCTGGNTAC 192

RESULT 15
US-09-880-107-1230/c
; Sequence 1230, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 1230
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA460665
US-09-880-107-1230

Query Match      1.6%; Score 36.4; DB 10; Length 293;
Best Local Similarity 61.7%; Pred. No. 1.8;
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2185 CCAGGCACCCAGATGATTCCTATGTCACCGACGAGAAATGGCAGCTCTTAAAGGAGA 2244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 94  CCACCTGCCCCCTCTCTTTAAGAAGATGGGGAGAAACTTCCAAATGTTTATAAGA 35

QY 2245 AAATCTTTGAGCCCCCAAAAAAAAAAAAAAAAAA 2278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 34  AAGTCTGTCAACACAAAAAAAAAAAAAAAAAAAAA 1

Search completed: February 22, 2003, 14:38:46
Job time : 330 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2003, 08:17:37 ; Search time 5848 Seconds
(without alignments)
11341.535 Million cell updates/sec

Title: US-10-024-197-24

Perfect score: 2279

Sequence: 1 agctaagcaggtacctgtca.....aaaaaaaaaaaaaaaaaaaaa 2279

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pt:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_in:*

18: em_nu:*

19: em_om:*

20: em_ov:*

21: em_ph:*

22: em_pl:*

23: em_ro:*

24: em_sts:*

25: em_un:*

26: em_vl:*

27: em_ba:*

28: em_htg_hum:*

29: em_htg_inv:*

30: em_htg_inv:*

31: em_htg_inv:*

32: em_htg_inv:*

33: em_htg_inv:*

34: em_htg_inv:*

35: em_htg_inv:*

36: em_htg_inv:*

37: em_htg_inv:*

38: em_htg_inv:*

39: em_htg_inv:*

40: em_htg_inv:*

41: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2279	100.0	2279	9	BC003356	BC003356 Homo sapi
2	2245.4	98.5	2259	9	HUMGCA	D13286 Human mRNA
3	2238	98.2	2275	9	HUMGCBPRC	M19285 Human gluco
4	2220	97.4	2587	9	HUMGCB	M16328 Human gluco
5	1785.6	78.4	1792	9	HUMGCB	K02920 Human lysos
6	1640.2	72.0	1661	6	I09351	I09351 Sequence 1
7	1587	69.6	1611	6	AX147658	AX147658 Sequence
8	1547.8	67.9	1551	6	AX147656	AX147656 Sequence
9	1544.2	67.8	1823	9	HUMGCB	D13287 Human mRNA
10	1544	67.7	1593	6	AX299779	AX299779 Sequence
11	1543	67.7	1551	6	AX191972	AX191972 Sequence
12	1543	67.7	1551	6	AX349364	AX349364 Sequence
13	1529.6	67.1	1592	6	AX299777	AX299777 Sequence
14	1343	58.9	1300	9	BC000349	BC000349 Homo sapi
15	1287	56.3	2711	9	AK090514	AK090514 Homo sapi
16	1172.4	51.4	1964	10	BC006663	BC006663 Mus muscu
17	1167	51.2	1884	10	MUSGCB	M24119 Mouse gluco
18	667	29.3	8850	9	HUMGCB1	J03059 Human gluco
19	667	29.3	75270	9	AF023268	AF023268 Homo sapi
20	667	29.3	178229	2	AL713999	AL713999 Homo sapi
21	660.6	29.0	5420	9	AF267177	AF267177 Homo sapi
22	659	28.9	7128	9	AF285236	AF285236 Pan trogl
23	652.6	28.6	5498	9	AF272641	AF272641 Gorilla g
24	639.4	28.1	5769	6	AX335078	AX335078 Sequence
25	639.4	28.1	5769	9	HUMGCB2	J03060 Human gluco
26	620.6	27.2	5494	9	AF272642	AF272642 Pan trogl
27	529.4	23.2	543	11	G06424	G06424 human STS W
28	180.2	7.9	16507	10	AY115108	AY115108 Mus muscu
29	180.2	7.9	210496	10	AC104327	AC104327 Mus muscu
30	171	7.5	188	9	HUMGCB1	M22212 Human gluco
31	163.4	7.2	184	9	HUMGCBGD	M20282 Human lysos
32	160.4	7.0	162	9	HUMGCBDA	M20248 Human gluco
33	141.8	6.2	345	4	AF134220	AF134220 Equus cab
34	130.4	5.7	138	6	AX341598	AX341598 Sequence
35	127.2	5.6	403	9	HUMGCBPRA	M18916 Human gluco
36	117.6	5.2	403	9	HUMGCBPRA	M18916 Human gluco
37	108.4	4.8	1949	6	AF352032	AF352032 Xanthomon
38	102.4	4.5	11594	1	AE012205	AE012205 Xanthomon
39	98.8	4.3	1785	3	AY060718	AY060718 Drosophil
40	92.2	4.0	9661	1	AE013177	AE013177 Thermoana
41	91.8	4.0	10007	1	AE011746	AE011746 Xanthomon
42	84.4	3.7	95	9	S44217	S44217 beta-glucos
43	82.8	3.6	95	9	S44219	S44219 beta-glucos
44	80.6	3.5	36958	3	CEFL1E6	Z81058 Caenorhabdi
45	79.6	3.5	95	9	S44552	S44552 beta-glucos

ALIGNMENTS

RESULT 1	BC003356	2279 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS	BC003356	Homo sapiens, glucosidase, beta; acid (includes glucosylceramidase), clone MGC:5191 IMAGE:2899915, mRNA, complete cds.			
DEFINITION	BC003356				
VERSION	BC003356.1	GI:13097170			
KEYWORDS	MGC.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 2279) Straussberg,R.				

TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 3 Row: g Column: 5.

FEATURES

source

1..2279

/organism="Homo sapiens"

/db_xref="LocusID:2629"

/db_xref="taxon:9606"

/clone="MGC:5191 IMAGE:2899915"

/tissue_type="Placenta, choriocarcinoma"

/clone_lib="NIH_MGC_10"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

106..1716

/codon_start=1

/product="glucosidase, beta; acid (includes

glucosylceramidase)

/protein_id="AAH03356.1"

/db_xref="GI:13097171"

/translation="MEFSSPSRECPKPLSRVSMAGSLGLLLQAVSWAGARPCI
PKRGYSVVCVNCATVDSDFDPFPALGTSRYSTRGRMELSMGPLOAHHTGT
GLLTLOPEOKFOVKVGGAMTDAAALNIALSPPAQNLKLSYSEEGIGYNIIRV
PHASCDPSIRTYADTPDQPLHNFSLPEEDTKLPIHLRALOLAQPVSLASPW
TSPLWLTNAYNGKSLKGPQGDIIHQWTWAFKLDAYAEHLQAFWATVAREPS
AGLSGYPFOCLGTPHEQDFIARDLPTLANSHTNRLMDLQDRLLLPWHAKVY
LTDPEAAKYVHGIAHYHLDLPAKATLGLERFLPNTLFASEACVSGKFWQSVYR
LGSWDRGMQYSHSIITNLLXHVVGWTDNLALNDEGPNWVRNFVDSPIIVDKDTF
YKOPMFHLGHFSKFIPEGSORVGLVASQKNDLDAVALMHPDGSVVVVVNLNRSSKDV
LTIKDPAVGFLFTISPCYSIHLYLRQ"

538 a 646 c 598 g 497 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 2279; DB 9; Length 2279;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCTAAGCAGGTACTGATCCTTGTGTTTGTAGTGATGCTCTATCCTTCAGAGAC 60
|||||
Db 1 AGCTAAGCAGGTACTGATCCTTGTGTTTGTAGTGATGCTCTATCCTTCAGAGAC 60
|||||
QY 61 TCTGAACCCCTGTGTTCTCTCATATGACCTGAGGGATGAGCTTTTCAAGT 120
|||||
Db 61 TCTGAACCCCTGTGTTCTCTCATATGACCTGAGGGATGAGCTTTTCAAGT 120
|||||
QY 121 CCTTCCAGAGAGATGTCCCAAGCCTTTAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180
|||||
Db 121 CCTTCCAGAGAGATGTCCCAAGCCTTTAGTAGGGTAAGCATCATGGCTGGCAGCCTC 180
|||||
QY 181 ACAGGATGCTTCTACTTCAGCAGTGTGCTGGGATCAGGTGCGCCCTGATCCCT 240
|||||
Db 181 ACAGGATGCTTCTACTTCAGCAGTGTGCTGGGATCAGGTGCGCCCTGATCCCT 240
|||||
QY 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATCTGTGACTCCTTT 300
|||||

Db 241 AAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGCAATGCCACATCTGTGACTCCTTT 300
|||||
QY 301 GACCCCCCGACCTTCTCTGCTGCTTGGTACCTTCAGCGCTATGAGAGTACACGACGTGGG 360
|||||
Db 301 GACCCCCCGACCTTCTCTGCTGCTTGGTACCTTCAGCGCTATGAGAGTACACGACGTGGG 360
|||||
QY 361 CGACGGATGAGCTGATATGGGCCCCATCCAGGCTAATACACGCGGCACAGGCTGCTA 420
|||||
Db 361 CGACGGATGAGCTGATATGGGCCCCATCCAGGCTAATACACGCGGCACAGGCTGCTA 420
|||||
QY 421 CTGACCTTCAGCAGCAGACAGAGTTCCAGAAAGTGAAGGATTTGGAGGGGCGCATGACA 480
|||||
Db 421 CTGACCTTCAGCAGCAGACAGAGTTCCAGAAAGTGAAGGATTTGGAGGGGCGCATGACA 480
|||||
QY 481 GATGCTGCTCTCTCAACATCTTGCCTGTACCCCTGCCCAAAATTTGCTACTTAA 540
|||||
Db 481 GATGCTGCTCTCTCAACATCTTGCCTGTACCCCTGCCCAAAATTTGCTACTTAA 540
|||||
QY 541 TCGTACTTCTCTGAAGAAGATCGGATATACATCATCGGCTACCCATGCGCAGCTGT 600
|||||
Db 541 TCGTACTTCTCTGAAGAAGATCGGATATACATCATCGGCTACCCATGCGCAGCTGT 600
|||||
QY 601 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 660
|||||
Db 601 GACTTCTCCATCCGACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 660
|||||
QY 661 TTGAGCTTCCAGAGGAGATACCAAGCTCAAGATACCCCTGATTCACGAGCCCTGCAG 720
|||||
Db 661 TTGAGCTTCCAGAGGAGATACCAAGCTCAAGATACCCCTGATTCACGAGCCCTGCAG 720
|||||
QY 721 TTGGCCAGCGTCCGCTTCTCACTCTTCCAGCCCTGGACATCACCCACTTGGCTCAAG 780
|||||
Db 721 TTGGCCAGCGTCCGCTTCTCACTCTTCCAGCCCTGGACATCACCCACTTGGCTCAAG 780
|||||
QY 781 ACCAATGAGCGGTGATGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
|||||
Db 781 ACCAATGAGCGGTGATGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCAC 840
|||||
QY 841 CAGACCTTGGCCAGATACTTTTGAAGTTCTTGGATGCTTATCTGAGCACAAGTTTACAG 900
|||||
Db 841 CAGACCTTGGCCAGATACTTTTGAAGTTCTTGGATGCTTATCTGAGCACAAGTTTACAG 900
|||||
QY 901 TTCTGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTC 960
|||||
Db 901 TTCTGGCAGTGACAGCTGAAATGAGCCTTCTGCTGGGCTGTGAGTGGATACCCCTTC 960
|||||
QY 961 CAGTGCCTGGGCTTACCCCTGAAACATCAGCAGACTTCAATGCCGCTGACCTAGTGCCT 1020
|||||
Db 961 CAGTGCCTGGGCTTACCCCTGAAACATCAGCAGACTTCAATGCCGCTGACCTAGTGCCT 1020
|||||
QY 1021 ACCCTGCCAACAGTACTACCAATGTCCGCTTCTCATGCTGGATGACCAACGCTTG 1080
|||||
Db 1021 ACCCTGCCAACAGTACTACCAATGTCCGCTTCTCATGCTGGATGACCAACGCTTG 1080
|||||
QY 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAACAGCTAAATATGTTAT 1140
|||||
Db 1081 CTGCTGCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAACAGCTAAATATGTTAT 1140
|||||
QY 1141 GGCAATGCTGACATGTTGTTACCTGGACTTTCGGCTCCAGCCAAAGCCCTTAGGGAG 1200
|||||
Db 1141 GGCAATGCTGACATGTTGTTACCTGGACTTTCGGCTCCAGCCAAAGCCCTTAGGGAG 1200
|||||
QY 1201 ACACACCGCTGTTCGCCAACACCATGCTCTTTGCCCTCAGAGGCTGTGTGGGCTCCAAG 1260
|||||
Db 1201 ACACACCGCTGTTCGCCAACACCATGCTCTTTGCCCTCAGAGGCTGTGTGGGCTCCAAG 1260
|||||
QY 1261 TTCTGGAGCAGAGTGTGGGCTAGGCTCTCTGGGATCGAGGATGAGATACAGCCACAGC 1320
|||||
Db 1261 TTCTGGAGCAGAGTGTGGGCTAGGCTCTCTGGGATCGAGGATGAGATACAGCCACAGC 1320
|||||
QY 1321 ATCATACAGAACTCTCTGTACCATGTGCTGGCTGGACGACCTTGGCCTGAAC 1380
|||||

QY 361 CGACGATGAGCTGAGTATGGGGCCCATCCAGGCTAATCAACGGGACAGGCGCTGCTA 420
Db 361 CGACGATGAGCTGAGTATGGGGCCCATCCAGGCTAATCAACGGGACAGGCGCTGCTA 420
QY 421 CTGACCTCCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTGGAGGGCCATGACA 480
Db 421 CTGACCTCCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTGGAGGGCCATGACA 480
QY 481 GATGCTGCTCTCAACATCTTGGCCCTGTCCACCCCTGCCAAAATTTGCTACTTAA 540
Db 481 GATGCTGCTCTCAACATCTTGGCCCTGTCCACCCCTGCCAAAATTTGCTACTTAA 540
QY 541 TGTGCTCTCTGAGAGGAATTCGATATACATCATCCGGGTACCATGGCCAGCTGT 600
Db 541 TGTGCTCTCTGAGAGGAATTCGATATACATCATCCGGGTACCATGGCCAGCTGT 600
QY 601 GACTTCTCCATCCGACCTTACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 660
Db 601 GACTTCTCCATCCGACCTTACCTATGACAGACACCCCTGATGATTTCCAGTTGCACAC 660
QY 661 TTCAGCTCCAGAGGAATACCAAGTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
Db 661 TTCAGCTCCAGAGGAATACCAAGTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
QY 721 TTGGCCCAAGCTCCGTTTCACTCTCTTGGCCAGCCCTGGACATACCCCTTGGCTCAAG 780
Db 721 TTGGCCCAAGCTCCGTTTCACTCTCTTGGCCAGCCCTGGACATACCCCTTGGCTCAAG 780
QY 781 ACCAATGGAGCGGTGAATGGAGGGGTACCTCAAGGACACCCCGGAGACATCTACCAC 840
Db 781 ACCAATGGAGCGGTGAATGGAGGGGTACCTCAAGGACACCCCGGAGACATCTACCAC 840
QY 841 CAGACCTGGCCAGATACCTTTGTAAGTTCCTGATGCTATGCTGAGCAAGATTTACAG 900
Db 841 CAGACCTGGCCAGATACCTTTGTAAGTTCCTGATGCTATGCTGAGCAAGATTTACAG 900
QY 901 TTCTGGGAGTACAGCTGAAATGAGCTTCTGCTGGGCTTTCAGTGGATACCCCTTC 960
Db 901 TTCTGGGAGTACAGCTGAAATGAGCTTCTGCTGGGCTTTCAGTGGATACCCCTTC 960
QY 961 CAGTGTCTGGGCTTACCCCTTCAACATCAGCGAGACTTCATTGCCCTGACCTAGTCTCT 1020
Db 961 CAGTGTCTGGGCTTACCCCTTCAACATCAGCGAGACTTCATTGCCCTGACCTAGTCTCT 1020
QY 1021 ACCCTGGCAACAGTACTCACCAATGTCCGCTTACTGCTGATGCTGATGACCAAGCTTG 1080
Db 1021 ACCCTGGCAACAGTACTCACCAATGTCCGCTTACTGCTGATGCTGATGACCAAGCTTG 1080
QY 1081 CTGCTGCTCCCTTGGCAAGTGGTACTGACAGACCCAGAGAGCTTAAATATGTTTCA 1140
Db 1081 CTGCTGCTCCCTTGGCAAGTGGTACTGACAGACCCAGAGAGCTTAAATATGTTTCA 1140
QY 1141 GGCATTGCTGTACATTGGTACCTGGACTTTCGCTTCCAGCCAAAGCCACCTAGGGAG 1200
Db 1141 GGCATTGCTGTACATTGGTACCTGGACTTTCGCTTCCAGCCAAAGCCACCTAGGGAG 1200
QY 1201 ACACACCGCTCTTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTTCCAAG 1260
Db 1201 ACACACCGCTCTTCCCAACACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTTCCAAG 1260
QY 1261 TTCTGGGAGCAGATGTGGGCTAGGCTTCCGCTTCCAGAGGCTGTGTGGGCTTCCAAG 1320
Db 1261 TTCTGGGAGCAGATGTGGGCTAGGCTTCCGCTTCCAGAGGCTGTGTGGGCTTCCAAG 1320
QY 1321 ATCATCACCAAGCTCTGTACCATGTGGTGGCTGACGACTGGAACCTGCCCTGAAC 1380
Db 1321 ATCATCACCAAGCTCTGTACCATGTGGTGGCTGACGACTGGAACCTGCCCTGAAC 1380
QY 1381 CCCGAAGGAGGACCAATTTGGTGGCTTAATTTGTGCGAGTCCCATCATTTGTAGACATC 1440
Db 1381 CCCGAAGGAGGACCAATTTGGTGGCTTAATTTGTGCGAGTCCCATCATTTGTAGACATC 1440

QY 1441 ACCAAGGACACGCTTTTACAAACAGCCCATGTTTACACACCTTGGCCACTTCAGCAAGTTC 1500
Db 1441 ACCAAGGACACGCTTTTACAAACAGCCCATGTTTACACACCTTGGCCACTTCAGCAAGTTC 1500
QY 1501 ATTCTCTGAGGCTCCAGAGAGTGGGCTGTTTGCAGTTCAGAAAGACGACCTTGGAGCGCA 1560
Db 1501 ATTCTCTGAGGCTCCAGAGAGTGGGCTGTTTGCAGTTCAGAAAGACGACCTTGGAGCGCA 1560
QY 1561 GTGCACTGATGATCCGATCCGATGCTGCTGTTTGTGTGCTTAAACCGCTCTCTTAAG 1620
Db 1561 GTGCACTGATGATCCGATCCGATGCTGCTGTTTGTGTGCTTAAACCGCTCTCTTAAG 1620
QY 1621 GATGTGCTCTTACCATCAAGGATCCCTGCTGTTGCTGCTTGGAGACAATCTCACCTGGC 1680
Db 1621 GATGTGCTCTTACCATCAAGGATCCCTGCTGTTGCTGCTTGGAGACAATCTCACCTGGC 1680
QY 1681 TACTCCATTCACACTTACCTGTGGCTGCGCAGTGTGGAGCAGATACTCAAGGAGGCAC 1740
Db 1681 TACTCCATTCACACTTACCTGTGGCTGCGCAGTGTGGAGCAGATACTCAAGGAGGCAC 1740
QY 1741 TGGCTCAGCTGGGCTTAAAGGACAGAGTCACTCAGCTCAGCTGCTGCTGTGACTAAAGA 1800
Db 1741 TGGCTCAGCTGGGCTTAAAGGACAGAGTCACTCAGCTCAGCTGCTGCTGTGACTAAAGA 1800
QY 1801 GGGCAGCAGGCGGCTGAGCTTACAGGAGCTTAAGCCAGGGGCAATGTTTGGGT 1860
Db 1801 GGGCAGCAGGCGGCTGAGCTTACAGGAGCTTAAGCCAGGGGCAATGTTTGGGT 1860
QY 1861 GACTCACTTTCCTCTAGGTGTG - CCAGGGCTGAGGCGCTAGAAAAGATCAGTA 1919
Db 1861 GACTCACTTTCCTCTAGGTGTG - CCAGGGCTGAGGCGCTAGAAAAGATCAGTA 1920
QY 1920 AGCCCCAGTGTCCCCCAGCCCCCATCTTATGTGAACATCGCTGTGCTGCTGCTT 1979
Db 1920 AGCCCCAGTGTCCCCCAGCCCCCATCTTATGTGAACATCGCTGTGCTGCTGCTT 1980
QY 1980 TGGAACTGGGCTGGCTCCAGGCTAGGCTAGGCTCAGTGTCCGTACAAACAGATC 2039
Db 1980 TGGAACTGGGCTGGCTCCAGGCTAGGCTAGGCTCAGTGTCCGTACAAACAGATC 2040
QY 2040 AGGCTGAGGCTAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCCAACTGGAGACTGTT 2099
Db 2040 AGGCTGAGGCTAAGGAAAAGAGAGACTAGGAAAGCTGGGCCCCAACTGGAGACTGTT 2100
QY 2100 TGTCTTCTGAGATGCAGAACTGGGCCCCGTGGAGCAGCTGTGATCATGAGGCGGA 2159
Db 2100 TGTCTTCTGAGATGCAGAACTGGGCCCCGTGGAGCAGCTGTGATCATGAGGCGGA 2160
QY 2160 AGCCTTAAAGCAGCAGGCTGTGCCAGCCAGCCAGATGCTTATGTCACAGCCAG 2219
Db 2160 AGCCTTAAAGCAGCAGGCTGTGCCAGCCAGCCAGATGCTTATGTCACAGCCAG 2220
QY 2220 GAAAAATGGCAGCTTCTTAAAGGAGAAAATGTTTGGAGCCC 2258
Db 2220 GAAAAATGGCAGCTTCTTAAAGGAGAAAATGTTTGGAGCCC 2259

RESULT 3

HUMGBPRC

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

HUMGBPRC 2275 bp mRNA linear PRI 27-APR-1993
Human glucocerebrosidase mRNA, complete cds.

M19285

M19285.1 GI:183027

glucocerebrosidase

Human erythroid cell line K562, cDNA to mRNA, clone p7-2-1.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2275)

Horowitz, M.

Unpublished (1988)

2 (bases 1 to 140)

Reiner, O., Wigderson, M. and Horowitz, M.

```

TITLE      Structural analysis of the human glucocerebrosidase genes
JOURNAL    DNA 7 (2), 107-116 (1988)
MEDLINE    88195776
PUBMED     3359914
COMMENT     Draft entry and computer-readable sequence for [2], [1] kindly
            provided by M. Horowitz, 13-MAY-1988.
            An alternative initiation codon is located at positions 193-185 and
            alternative polyadenylation signals at positions 2250-2255 (major)
            and 1774-1779 (minor).
FEATURES   Location/Qualifiers
            source          1..2275
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        <1..2275
                        /product="GCB mRNA"
                        123..1733
                        /note="glucocerebrosidase"
                        /codon_start=1
                        /protein_id="AAA35880.1"
                        /db_xref="GI:183028"
                        /translation="MEFSSPSRECPKPLSRYSIMAGSLTGILLIQAVSWASGARPCI
                        PKFGYSVVCVATYCDSEDPPTPALGTFPSRYSRSGRMELSMGPLOAHNGT
                        GLLTLQPEORFQVKGFGGAMTDAAALILALSPALNLLKSVFSEGGIYNIIV
                        PWASCDISIRTYADYDPDFQNLNFIPEEDTKLILPHRALOLALORVSLIASEPW
                        TSPWLKNTGAVNGKSGIKGORDIYHOFMARFYKFLDVAEHLKLOFWAVTAENFES
                        AGLLSGYPFQGLGTBEHORDFYAROLGTLANSIHNNVRLMLDDQLLLPHWAKVY
                        LNDPRAKYVGIAYHYLOFLAPARATLGETHRFLFPNFMIFASACVGFWEQSVR
                        LGSMDRCMQYSHSIIITLLHYVGTDMNIALNPEGGPNWVRNEFSDPIIVDITKDF
                        YKQPMFLGHFSRFPEGSQRVCLVASQNDLDAVALMHPDGSVAVVVLNRSKQVP
                        LTIKDPVGLFETISPGYSIHTIIMHQ*
BASE COUNT 523 a 652 c 596 g 503 t
ORIGIN      32 bp upstream of KpnI site; chromosome 1q21-31.
Query Match 98.2%; Score 2238; DB 9; Length 2275;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2252; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1 AGCTAAGGAGGAGTACCTGATCCTTGTGTTTGTGTTAGTGGATCCTCTATCTCTTCAGAGAC 60
Db 18 AGCTAAGGAGGAGTACCTGATCCTTGTGTTTGTGTTAGTGGATCCTCTATCTCTTCAGAGAC 77
Qy 61 TCTGGAACCCCTGTGCTCTCTCTTCATCTAATGACCCCTCAGGGGATGAGTTCATCAAGT 120
Db 78 TCTGGAACCCCTGTGCTCTCTCTTCATCTAATGACCCCTCAGGGGATGAGTTCATCAAGT 137
Qy 121 CCTTCCAGAGAGAAATGCCAAGCCTTGTAGTAGGTAAGCATGATGCTCGCAGCCCTC 180
Db 138 CCTTCCAGAGAGAAATGCCAAGCCTTGTAGTAGGTAAGCATGATGCTCGCAGCCCTC 197
Qy 181 ACAGGATGCTTACTTACGCGAGTGTGTTGGCATCAGGTGCCGCCCTCGCATCCCT 240
Db 198 ACAGGATGCTTACTTACGCGAGTGTGTTGGCATCAGGTGCCGCCCTCGCATCCCT 257
Qy 241 AAAAGCTCGGCTACAGCTCGGTGTGTGTCTGCAATGCCACATCTGTGACTCCCTT 300
Db 258 AAAAGCTCGGCTACAGCTCGGTGTGTGTCTGCAATGCCACATCTGTGACTCCCTT 317
Qy 301 GACCCCGGACCTTTCTGCGCCTTGGTACCTTCAGCCGCTATGAGAGTACAGCGAGTGGG 360
Db 318 GACCCCGGACCTTTCTGCGCCTTGGTACCTTCAGCCGCTATGAGAGTACAGCGAGTGGG 377
Qy 361 CGACGATGAGCTGAGTATGGGCGCCATCCAGGCTAATCACAGGGCACAGGCCCTGCTA 420
Db 378 CGACGATGAGCTGAGTATGGGCGCCATCCAGGCTAATCACAGGGCACAGGCCCTGCTA 437
Qy 421 CTGACCTCAGCAGACAGAAATTCAGAAATGAAAGGATTTGGAGGGCCATGACA 480
Db 438 CTGACCTCAGCAGACAGAAATTCAGAAATGAAAGGATTTGGAGGGCCATGACA 497
Qy 481 GATGCTGCTGCTCTCAACATCTTGCCTGTACCCCTGTACCCCTGTACCTCTAA 540
Db 498 GATGCTGCTGCTCTCAACATCTTGCCTGTACCCCTGTACCCCTGTACCTCTAA 557

```


QY	558	AGGAATCGGATATAACATCATCCGGGTACCATTGCCAGCTGTGACTTCTCCATCCGCAC	617
DB	888	AGGAATCGGATATAACATCATCCGGGTACCATTGCCAGCTGTGACTTCTCCATCCGCAC	947
QY	618	CTACACCTATGCGAGACACCCCTGATGATTTCCAGTTTGCACACTTTCAGCCTCCGAGAGGA	677
DB	948	CTACACCTATGCGAGACACCCCTGATGATTTCCAGTTTGCACACTTTCAGCCTCCGAGAGGA	1007
QY	678	AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTTGGCCCGACGCTCCCGT	737
DB	1008	AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTTGGCCCGACGCTCCCGT	1067
QY	738	TTCACTTCCTTGCCAGCCCTTGGACATCACCCACTTTGGCTCAAGACCAATTGGAGCGGTGAA	797
DB	1068	TTCACTTCCTTGCCAGCCCTTGGACATCACCCACTTTGGCTCAAGACCAATTGGAGCGGTGAA	1127
QY	798	TGGGAAGGGGTCACTCAAGGAGCAGCCGCGAGACATCTACACACAGCACTCTGGCCGACAGTA	857
DB	1128	TGGGAAGGGGTCACTCAAGGAGCAGCCGCGAGACATCTACACACAGCACTCTGGCCGACAGTA	1187
QY	858	CTTTGTGGAAGTTCTTGATGCCATATGCTGAGCACAAAGTTACAGTTCTGGGCAAGTGACAGC	917
DB	1188	CTTTGTGGAAGTTCTTGATGCCATATGCTGAGCACAAAGTTACAGTTCTGGGCAAGTGACAGC	1247
QY	918	TGAAATGAGCTTCTGCTGGGCTGTGTAGTGGATACCCCTTCCAGTGCTGGGCTTCAC	977
DB	1248	TGAAATGAGCTTCTGCTGGGCTGTGTAGTGGATACCCCTTCCAGTGCTGGGCTTCAC	1307
QY	978	CCCTGAACATCAGCGAGACTTCAATGCGCGTGACCTTAGTGCTTACCCCTCGCCACAGTAC	1037
DB	1308	CCCTGAACATCAGCGAGACTTCAATGCGCGTGACCTTAGTGCTTACCCCTCGCCACAGTAC	1367
QY	1038	TCACCACAATGTCCGCTACTCATGTGTGATGACCAACGCTTGTGCTGCCCCACTGGGC	1097
DB	1368	TCACCACAATGTCCGCTACTCATGTGTGATGACCAACGCTTGTGCTGCCCCACTGGGC	1427
QY	1098	AAAGTGTGTACTGACAGACCCAGAAGCACTAAATATGTTTCATGGCATTGTGTACATTG	1157
DB	1428	AAAGTGTGTACTGACAGACCCAGAAGCACTAAATATGTTTCATGGCATTGTGTACATTG	1487
QY	1158	GTACCTGGACTTTCTGGCTCCAGGCCAAAGCCACCTTAGGGGAGACACACCGCCTGTTCC	1217
DB	1488	GTACCTGGACTTTCTGGCTCCAGGCCAAAGCCACCTTAGGGGAGACACACCGCCTGTTCC	1547
QY	1218	CAACACCATGCTTTGCGCTCAGAGCCCTGTGGGCTCCAGTTCTGGGAGCAGAGTGT	1277
DB	1548	CAACACCATGCTTTGCGCTCAGAGCCCTGTGGGCTCCAGTTCTGGGAGCAGAGTGT	1607
QY	1278	GGGCTAGGCTTCCTGGATGCGAGGATGCGATCAGCCACAGCATCATCAGAACCTTCCT	1337
DB	1608	GGGCTAGGCTTCCTGGATGCGAGGATGCGATCAGCCACAGCATCATCAGAACCTTCCT	1667
QY	1338	GTACCATGTGTGGCTTGGACCGACTTGAACCTTTGGCCTTGAACCCCGAAGGAGGACCCAA	1397
DB	1668	GTACCATGTGTGGCTTGGACCGACTTGAACCTTTGGCCTTGAACCCCGAAGGAGGACCCAA	1727
QY	1398	TTGGGTGGCTTAACCTTTGCGACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTTTA	1457
DB	1728	TTGGGTGGCTTAACCTTTGCGACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTTTA	1787
QY	1458	CAAAACAGCCCATGTTCTACACCTTTGGCCACTTTCAGCAGTTTCATCTCTGAGGCTCCCA	1517
DB	1788	CAAAACAGCCCATGTTCTACACCTTTGGCCACTTTCAGCAGTTTCATCTCTGAGGCTCCCA	1847
QY	1518	GAGATGGGGCTGGTTGCCAGTCAAGAAGACGACCTTGGAGCGAGTGGCACTGATGCATCC	1577
DB	1848	GAGATGGGGCTGGTTGCCAGTCAAGAAGACGACCTTGGAGCGAGTGGCACTGATGCATCC	1907
QY	1578	CGATGGCTTCGCTGTGTGGTCTGCTAAACCCGCTCCTCTTAAGAGTATGTGCCTCTTACCAT	1637
DB	1908	CGATGGCTTCGCTGTGTGGTCTGCTAAACCCGCTCCTCTTAAGAGTATGTGCCTCTTACCAT	1967

QY	1638	CAAGGATCTGCTGGGCTTCTCGAGACAAATCTCACTGGCTACTTCATTCACACCTA	1697
DB	1968	CAAGATCTTCGCTGGGCTTCTCGAGACAAATCTCACTGGCTACTTCATTCACACCTA	2027
QY	1698	CCCTGTGGCTGCGCCAGTGAATGAGCAGATATCAAGGAGGCACTGGGCTCAGCCTGGGCA	1757
DB	2028	CCCTGTGGCTGCGCCAGTGAATGAGCAGATATCAAGGAGGCACTGGGCTCAGCCTGGGCA	2087
QY	1758	TTAAAGGACAGAGTCAGCTCACGCTGTCTGTGACTAAAGAGGCGACAGCAGGGCCAG	1817
DB	2088	TTAAAGGACAGAGTCAGCTCACGCTGTCTGTGACTAAAGAGGCGACAGCAGGGCCAG	2147
QY	1818	TGTGAGCTTACAGCGACGTAAAGCCAGGGGCAATGGTTTGGGTGACTCACTTTCCCTCT	1877
DB	2148	TGTGAGCTTACAGCGACGTAAAGCCAGGGGCAATGGTTTGGGTGACTCACTTTCCCTCT	2207
QY	1878	AGTGTGTGCCAGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCGCCAGTGTCCGCCCA	1937
DB	2208	AGTGTGTGCCAGGGCTGGAGGCCCTAGAAAAAGATCAGTAAGCCGCCAGTGTCCGCCCA	2267
QY	1938	GCCTCATGCTTATGTGAACATGCGCTGTGTCGTCCTTTTGGAACTGGGCTCGGCT	1997
DB	2268	GCCTCATGCTTATGTGAACATGCGCTGTGTCGTCCTTTTGGAACTGGGCTCGGCT	2326
QY	1998	CCAGGCTTAGGTGTAGCTCACTGTCCGTACAAACACAAAGATCAGGCTCAGGATGAAGAA	2057
DB	2327	CCAGGCTTAGGTGTAGCTCACTGTCCGTACAAACACAAAGATCAGGCTCAGGATGAAGAA	2386
QY	2058	AACAAGAGACTAGGAAGCTGGGCGCCAAAACCTGGAGACTGTTCCTTCCTGAGATGC	2117
DB	2387	AAGAAGACTAGGAAGCTGGGCGCCAAAACCTGGAGACTGTTCCTTCCTGAGATGC	2446
QY	2118	AGAACTGGGCGCGTGGAGCAGCAGTGTCAATCAGGCGGGAAGCCCTTAAAGCAGCAGCG	2177
DB	2447	AGAACTGGGCGCGTGGAGCAGCAGTGTCAATCAGGCGGGAAGCCCTTAAAGCAGCAGCG	2506
QY	2178	GGTGTGCCAGGACCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTCTTA	2237
DB	2507	GGTGTGCCAGGACCCAGATGATTCCTATGGCACCAGCCAGGAAAAATGGCAGCTCTTA	2566
QY	2238	AAGGAGAAATGTTTGAGCCCC	2258
DB	2567	AAGGAGAAATGTTTGAGCCCC	2587
RESULT 5			
HUMGCBL		1792 bp	linear
LOCUS		Human lysosomal glucocerebrosidase mRNA, complete cds.	
DEFINITION			
ACCESSION		K02920	
VERSION		K02920.1	GI:183021
KEYWORDS		glucocerebrosidase.	
SOURCE		Human hepatoma, cDNA to mRNA, clone pGCL.	
ORGANISM		Homo sapiens	
REFERENCE		Emkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		Tsuiji, S., Choudary, P.V., Martin, B.M., Winfield, S., Barranger, J.A. and Ginn, E. I.	
TITLE		Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase	
JOURNAL		J Biol Chem. 261 (1), 50-53 (1986)	
MEDLINE		86085859	
PUBMED		3001061	
COMMENT		A preprint for [1] was kindly provided by E.I.Ginnis, 01-AUG-1985 Glucocerebrosidase is a membrane bound lysosomal enzyme, which catalyzes the hydrolysis of glucocerebrosidase. Allelic mutations in the gene coding for glucocerebrosidase cause the alteration of normal synthesis and/or processing, resulting in sphingolipidosis. The precise nature of the genetic changes resulting in the three phenotypes of this disease have not yet been determined. The glucocerebrosidase sequence reported by [1] differs by one AA from human placental glucocerebrosidase.	

FEATURES
source

Location/Qualifiers

1..1792
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>1792
/product="lrg mRNA"
46..1596
/note="lysosomal glucocerebrosidase precursor, EC 3.2.1.45"
/codon_start=1
/protein_id="AAA35877.1"
/db_xref="GI:183022"
/translation="MAGSLTGLLLQLQAVSWASGARPICPKSFYSSVVCNATYCDSDPFPALGTFSEVSTGRSRMELSMGPLOAHNTGTGLLLTLQPKQKVKVFGGAMTDAALNIALSPPAONLLKSYFEEGIGYNIIRVPMASCDSEIRTYTADTPDQFOLNFSPEDETKLIPHLRALQAPVSLASPTPTWLTNGAVNGKSLKGOPGDIYHOTWARYFVKFLDAYAEHLQFVAVTAENESAGLLSGYPQCGLTFPEHORDLIARDLGLTSHNRLLMDQDRLLLPHWAKVVLDPDEAAKTVHGIJAVHWYLDFLAPKATLGTHRLFTMLFASAEACQKFWQSVRLGSDWRGMYSHSIITNLLYHVGWTDNMLALNPEGGPNVVRNFVDPPIIVDITKDTFYKQPMFYHLGHFSKFIPEGSORGLVASOKNDLDAVALMHPDGSVAVVVLNRRSKDVPLTIKDPVAFGLETISPYSIHTYLRRO"
46..102
/note="lysosomal glucocerebrosidase signal peptide"
103..1593
/product="lysosomal glucocerebrosidase"
mat_peptide 406 a 528 c 466 g 392 t
BASE COUNT 311 bp upstream of PstI site; chromosome 1q21.

sig_peptide

mat_peptide

Query Match 78.4%; Score 1785.6; DB 9; Length 1792;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1788; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 121 CCTTCCAGAGAGATGTCCTCCAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTC 180
Db 1 CCTTCCAGAGAGATGTCCTCCAGCCTTTGAGTAGGTAAGCATCATGGCTGGCAGCCTC 60
QY 181 ACAGGATTCCTTACTTCAGGACGTCGTCGGGATCAGGTGCGCCGCCCTGCAATCCCT 240
Db 61 ACAGGATTCCTTACTTCAGGACGTCGTCGGGATCAGGTGCGCCGCCCTGCAATCCCT 120
QY 241 AAAAGCTTCGGCTACAGCTCGGTGTCGTCTGCAATGCCACATACGTGACCTCCCTTT 300
Db 121 AAAAGCTTCGGCTACAGCTCGGTGTCGTCTGCAATGCCACATACGTGACCTCCCTTT 180
QY 301 GACCCCGGACCTTCTCCTGCCCTTGTACCTTCAGCCGCTATGAGATACAGCAGTGGG 360
Db 181 GACCCCGGACCTTCTCCTGCCCTTGTACCTTCAGCCGCTATGAGATACAGCAGTGGG 240
QY 361 CGACGGATGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGGCAGCGCCTGCTA 420
Db 241 CGACGGATGAGCTGAGTATGGGGCCCATCCAGGCTAATCACAGGGCAGCGCCTGCTA 300
QY 421 CTCACCTCGCAGCCAGCAAGATTCAGAAAGTGAAGGGATTTGGAGGGCCCATGACA 480
Db 301 CTCACCTCGCAGCCAGCAAGATTCAGAAAGTGAAGGGATTTGGAGGGCCCATGACA 360
QY 481 GATGCTGCTCTCAACATCCCTTGGCCCTGTCAACCCCTGCCCCAAATTTGCTACTTAA 540
Db 361 GATGCTGCTCTCAACATCCCTTGGCCCTGTCAACCCCTGCCCCAAATTTGCTACTTAA 420
QY 541 TCCTACTTCTCAAGAGGAATCGGATATACATCATCCGGGTACCCATGCCAGCTGT 600
Db 421 TCCTACTTCTCAAGAGGAATCGGATATACATCATCCGGGTACCCATGCCAGCTGT 480
QY 601 GACTTCTCCATCCGACCTACACCTATGAGACACACCCCTGATGATTTCCAGTTGACAC 660
Db 481 GACTTCTCCATCCGACCTACACCTATGAGACACACCCCTGATGATTTCCAGTTGACAC 540
QY 661 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAG 720
Db 541 TTCAGCCTCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCACCGAGCTGCAG 600

QY 721 TTGGCCAGCGTCCCGTTTCACTCTTGGCAGCCCTGGACATCACCCACTTGGCTCAAG 780
Db 601 TTGGCCAGCGTCCCGTTTCACTCTTGGCAGCCCTGGACATCACCCACTTGGCTCAAG 660
QY 781 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGAGACATCTACCAC 840
Db 661 ACCAATGGAGCGGTGAATGGGAAGGGTCACTCAAGGGACAGCCCGAGACATCTACCAC 720
QY 841 CAGACCTGGGCGAGATACTTTGTGAAGTTCTTGGATGCCCTATCTGCTGAGCACAAGTTACAG 900
Db 721 CAGACCTGGGCGAGATACTTTGTGAAGTTCTTGGATGCCCTATCTGCTGAGCACAAGTTACAG 780
QY 901 TTCTGGCAGTGACAGCTGAAAATAGCCCTTCTGCTGGGCTGTGTAGTGGATACCCCTTC 960
Db 781 TTCTGGCAGTGACAGCTGAAAATAGCCCTTCTGCTGGGCTGTGTAGTGGATACCCCTTC 840
QY 961 CAGTGCCTGGGCTTCAACCCCTGAAACATCAGCAGAGCTTCAATGCCCGTGAACCTAGTTCCT 1020
Db 841 CAGTGCCTGGGCTTCAACCCCTGAAACATCAGCAGAGCTTCAATGCCCGTGAACCTAGTTCCT 900
QY 1021 ACCCTGCCAACAGTACTACCCACAAATGTCCGCTACTCATGCTGGATGACCAACGCTTG 1080
Db 901 ACCCTGCCAACAGTACTACCCACAAATGTCCGCTACTCATGCTGGATGACCAACGCTTG 960
QY 1081 CTGCTGCCCACTGGGCAAGGTGGTACTCAGACAGCCCAAGACAGCTAAATATGTTTCA 1140
Db 961 CTGCTGCCCACTGGGCAAGGTGGTACTCAGACAGCCCAAGACAGCTAAATATGTTTCA 1020
QY 1141 GGCATTGCTGTACATTTGCTGACTTCTGCTGCTCCAGCAAGCCACCTTAGGGGAG 1200
Db 1021 GGCATTGCTGTACATTTGCTGACTTCTGCTGCTCCAGCAAGCCACCTTAGGGGAG 1080
QY 1201 ACACACCGCTTGTCCCAACACCATGCTCTTTGCCCTCAGAGGCTGTGTGGGCTCCAAG 1260
Db 1081 ACACACCGCTTGTCCCAACACCATGCTCTTTGCCCTCAGAGGCTGTGTGGGCTCCAAG 1140
QY 1261 TTCTGGGAGCAGAGTGTCCGCTAGGCTCTGCTGGATCGAGGATGACAGTACAGCCACAG 1320
Db 1141 TTCTGGGAGCAGAGTGTCCGCTAGGCTCTGCTGGATCGAGGATGACAGTACAGCCACAG 1200
QY 1321 ATCATCAGAACCTCTGTACCATGTGTGGCTGGACGACCTGGAACCTTGGCCCTGAAC 1380
Db 1201 ATCATCAGAACCTCTGTACCATGTGTGGCTGGACGACCTGGAACCTTGGCCCTGAAC 1260
QY 1381 CCGAGGAGCAGCCCAATTTGGCTGCTAACTTTCTCAGCAGTCCCATCATTTGTAGACATC 1440
Db 1261 CCGAGGAGCAGCCCAATTTGGCTGCTAACTTTCTCAGCAGTCCCATCATTTGTAGACATC 1320
QY 1441 ACCAAGGACAGCTTTTACAAACAGCCCATGTTCTTACCACTTGGCCCTCAGCAAGTTC 1500
Db 1321 ACCAAGGACAGCTTTTACAAACAGCCCATGTTCTTACCACTTGGCCCTCAGCAAGTTC 1380
QY 1501 ATTCTGAGGCTCCCGAGAGAGTGGGCTGTTGCCAGTCAAGAAAGACGCTGGAGCGCA 1560
Db 1381 ATTCTGAGGCTCCCGAGAGAGTGGGCTGTTGCCAGTCAAGAAAGACGCTGGAGCGCA 1440
QY 1561 GTGGCTGATGATCCCGATGGCTCTGCTGTGTGGTGGTCTAAACCGCTCTCTTAAG 1620
Db 1441 GTGGCTGATGATCCCGATGGCTCTGCTGTGTGGTGGTCTAAACCGCTCTCTTAAG 1500
QY 1621 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTTGGAGACAATCTACCTGGC 1680
Db 1501 GATGTGCTCTTACCATCAAGGATCCTGCTGTGGGCTTCTTGGAGACAATCTACCTGGC 1560
QY 1681 TACTCCATTCACAGCTTACCTGTGGGCTGGCAGTGTGAGGACAGATCTCAAGAGGCAC 1740
Db 1561 TACTCCATTCACAGCTTACCTGTGGGCTGGCAGTGTGAGGACAGATCTCAAGAGGCAC 1620
QY 1741 TGGGCTCAGCCTGGGCATTTAAAGGACAGAGTCAAGCTCACACGCTGTCTGTGACTAAGA 1800
Db 1621 TGGGCTCAGCCTGGGCATTTAAAGGACAGAGTCAAGCTCACACGCTGTCTGTGACTAAGA 1680
QY 1801 GGGCAGCAGAGGGGCGAGTGTGAGCTTACAGCGAGTAAAGCCAGGGGCAATGGTTGGGT 1860


```

QY 1666 ACATCTACCTGGCTGCTACCTGATTCACACCTACCTGCTGGCGTGGCAGTGA 1716
DB 1561 ACAATCTACCTGGCTGCTACCTGATTCACACCTACCTGCTGGCGTGGCAGTGA 1611

RESULT 8
LOCUS AX147656 1551 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0136603.
ACCESSION AX147656
VERSION AX147656.1 GI:14346712
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.
TITLE Recombinant adeno-associated virus virions for the treatment of
lysosomal disorders
JOURNAL Patent: WO 0136603-A 5 25-MAY-2001;
Avigen, Inc. (US); Children's Hospital Medical Center of Northern
California (US)
FEATURES
source location/Qualifiers
1..1551
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1551
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:14346713"
/translation="MAGSLTGLLLQLQAVSWASGARCPKIPKFGYSVVVCNATYCDSD
FDPTPALGTFSYSTRSGRMELSMGPIQANHTGTGGLLTLQPQKPKVKGFGG
AMTDAALNLIALSPALNLLKSYSEEGYNIIRVPMASCDPSIRTYADTPDD
FOLHNFSLPEDEYKLPILIRALQAPSVLASPWTPTLKTNGVANGSLKG
OPGDYTHWAFKFLDIAEHLQFQWAVTAENEPAGLSGYPFQCLGFTPEHOR
FLIARDLPTLANSTHNVRLMLDQRLPLPHWAKVVLTPDEAAKVVYHGVYLD
FLAPAKATLGTHRLPNTMLFASAEACVGSKFWEOSVRLGSMWDMGYSHSILTNLLY
HVGWTDNLALNPEGGPNVRFVDSPLIVDITKDTFYKOPMEFYHLGHSKFIPIBS
ORGLVASQKNDLDAVALMHPDGSVAVVYLNSSKDVLTIKDPVAGFLETISPYSI
HTYLWHQ"
BASE COUNT 344 a 474 c 387 g 346 t
ORIGIN

Query Match 67.9%; Score 1547.8; DB 6; Length 1551;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCGTCACAGGATGCTTCTACTCAGGAGTGTGCGGATCAGGTGCC 225
DB 1 ATGGCTGGCAGCGTCACAGGATGCTTCTACTCAGGAGTGTGCGGATCAGGTGCC 60

QY 226 CGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTCTGTGCAATGCCACA 285
DB 61 CGCCCTGTCATCCCTAAAGCTTCGGCTACAGCTCGGTGGTGTGTGTGCAATGCCACA 120

QY 286 TACTGTGACTCTTTGACCCCGACCTTTCTCGCTTGGTACCTTTCAGCCGCTATGAG 345
DB 121 TACTGTGACTCTTTGACCCCGACCTTTCTCGCTTGGTACCTTTCAGCCGCTATGAG 180

QY 346 AGTACACGAGTGGCGAGGATGAGTGTGATGGGCGCCATCCAGCTAATCACACG 405
DB 181 AGTACACGAGTGGCGAGGATGAGTGTGATGGGCGCCATCCAGCTAATCACACG 240

QY 406 GGCACAGGCTGCTACTACCTGCTGCGGACGACAGAGTTCAGAAAGTGAAGGATTT 465
DB 241 GGCACAGGCTGCTACTACCTGCTGCGGACGACAGAGTTCAGAAAGTGAAGGATTT 300

QY 466 GGAGGGGCGCATGACAGATGCTGTGCTCTCAACATCTTGGCTGTGACCCCGTGCACAA 525
DB 301 GGAGGGGCGCATGACAGATGCTGTGCTCTCAACATCTTGGCTGTGACCCCGTGCACAA 360

```

```

QY 526 AATTTGCTACTTAAATCGTACTTCTCTGAGAGGAATCGGATATATACATCATCFCGGGTA 585
DB 361 AATTTGCTACTTAAATCGTACTTCTCTGAGAGGAATCGGATATATACATCATCFCGGGTA 420

QY 586 CCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGAT 645
DB 421 CCATGGCCAGCTGTGACTTCTCCATCCGCACCTACACCTATGCAGACACCCCTGATGAT 480

QY 646 TTCAGTTGGACAACTTCAGCTCCCGAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 705
DB 481 TTCAGTTGGACAACTTCAGCTCCCGAGAGGAAGATACCAAGCTCAAGATACCCCTGATT 540

QY 706 CACGAGCCCTGCAGTTGGCCAGCGTCCGCTTTCCTCTTGCAGGCCCTGGACATCA 765
DB 541 CACGAGCCCTGCAGTTGGCCAGCGTCCGCTTTCCTCTTGCAGGCCCTGGACATCA 600

QY 766 CCACCTTGGCTCAAGACCAATGAGCGGTGAATGGAGGGGTCACTCAAGGACAGGCC 825
DB 601 CCACCTTGGCTCAAGACCAATGAGCGGTGAATGGAGGGGTCACTCAAGGACAGGCC 660

QY 826 GGAGACATCTACCACAGACCTGGCCAGATACCTTGTGAAGTTCTCGATGCTTATGCT 885
DB 661 GGAGACATCTACCACAGACCTGGCCAGATACCTTGTGAAGTTCTCGATGCTTATGCT 720

QY 886 GAGCACAAGTTACAGTTCTGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 945
DB 721 GAGCACAAGTTACAGTTCTGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780

QY 946 AGTGATACCCCTTCCAGTGGCTTCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1005
DB 781 AGTGATACCCCTTCCAGTGGCTTCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 840

QY 1006 CGTACCTAGGTCTCTACCTCCGCAACAGTACTCACCACAACTGCGGCTTACTCATGCTG 1065
DB 841 CGTACCTAGGTCTCTACCTCCGCAACAGTACTCACCACAACTGCGGCTTACTCATGCTG 900

QY 1066 GATGACCAACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125
DB 901 GATGACCAACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

QY 1126 GCTAAATATCTTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185
DB 961 GCTAAATATCTTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

QY 1186 GCCACCTAGGAGAGACACACCGCTGTTCCCAACCAATGCTTCTTGGCTCAGAGGCC 1245
DB 1021 GCCACCTAGGAGAGACACACCGCTGTTCCCAACCAATGCTTCTTGGCTCAGAGGCC 1080

QY 1246 TGTGTGGCTTCCAAAGTTCTGGGAGCAGAGTGTGGGCTTAGGCTTCTTGGATCAGGAGTG 1305
DB 1081 TGTGTGGCTTCCAAAGTTCTGGGAGCAGAGTGTGGGCTTAGGCTTCTTGGATCAGGAGTG 1140

QY 1306 CAGTACAGCCACAGCATCATCAGAACCTTCTGTACCATGTGTGCTGCGCTGGACCGACTGG 1365
DB 1141 CAGTACAGCCACAGCATCATCAGAACCTTCTGTACCATGTGTGCTGCGCTGGACCGACTGG 1200

QY 1366 AACCTTCCCTGAAACCCCGAAGGAGGACCAATTTGGTGTGCTAACTTTGTCAGACAGTCCC 1425
DB 1201 AACCTTCCCTGAAACCCCGAAGGAGGACCAATTTGGTGTGCTAACTTTGTCAGACAGTCCC 1260

QY 1426 ATCAATTTAGACATACCAAGGACAGCTTTTACAAACAGCCCATGTTCTTACCACCTGGC 1485
DB 1261 ATCAATTTAGACATACCAAGGACAGCTTTTACAAACAGCCCATGTTCTTACCACCTGGC 1320

QY 1486 CACTTCAGCAAGTTTCACTCTGAGGCTCCAGAGAGTGGGGTGGTTCGCAAGTCAAGAG 1545
DB 1321 CACTTCAGCAAGTTTCACTCTGAGGCTCCAGAGAGTGGGGTGGTTCGCAAGTCAAGAG 1380

QY 1546 AACGACCTGGACAGCTGACATGATGATCCGATGGCTCTGCTGTGTGTGTGTGTGTGTGTGT 1605
DB 1381 AACGACCTGGACAGCTGACATGATGATCCGATGGCTCTGCTGTGTGTGTGTGTGTGTGTGT 1440

```

QY 1606 AACCGCTCTCTAAGAGTGTGCTCTTACCATCAAGAGTTCCTGCTGGGCTTCTCTGGAG 1665
 |||||||
 Db 1441 AACCGCTCTCTAAGAGTGTGCTCTTACCATCAAGAGTTCCTGCTGGGCTTCTCTGGAG 1500
 |||||||
 QY 1666 ACAACTCACCCTGGTACTCCTCAATTCACACCTACCTGCTGGCTGCGCAGTGA 1716
 |||||||
 Db 1501 ACAACTCACCCTGGTACTCCTCAATTCACACCTACCTGCTGGCTGCGCAGTGA 1551
 |||||||

RESULT 9
 HUNGCB
 LOCUS HUNGCB 1823 bp mRNA linear PRI 03-FEB-1999
 DEFINITION Human mRNA for alternative splicing product of glucocerebrosidase
 pseudogene, complete cds.
 ACCESSION D13287
 VERSION 1
 KEYWORDS alternative splicing product; glucocerebrosidase; pseudogene.
 SOURCE Homo sapiens (library: lambda hG12) cell-line HL-60 cDNA to mRNA,
 clone pGC12A.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1823)
 AUTHORS Imai,K., Nakamura,M., Yamada,M., Asano,A., Yokoyama,S., Tsuji,S.
 and Glns,E.I.
 TITLE A novel transcript from a pseudogene for human glucocerebrosidase
 in non-Gaucher disease cells
 JOURNAL Gene 136 (1-2), 365-368 (1993)
 MEDLINE 94124033
 REFERENCE 2 (bases 1 to 1823)
 AUTHORS Nakamura,M.
 TITLE Direct Submission
 JOURNAL Submitted (21-SEP-1992) Masahiko Nakamura, Osaka University,
 Institute for Protein Research, 3-2 Yamadaoka, Suita, Osaka 565,
 Japan (E-mail:masahiko@protein.osaka-u.ac.jp, Tel:06-879-8628,
 Fax:06-879-8629)
 COMMENT On Jun 5, 1994 this sequence version replaced gi:441043.
 FEATURES
 source
 1. 1823
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="HL-60"
 /clone_lib="lambda hG12"
 209..1000
 /note="alternative splicing product of glucocerebrosidase
 pseudogene"
 CDS
 /codon_start=1
 /product="glucocerebrosidase"
 /protein_id="BAA02546.1"
 /db_xref="gi:2160388"
 /translation="MASCDFSRITVTYADTDDFOLHNFSLPEEDTKLKIPLIHALQ
 LAQRPVSLASPTWSPRLKTRGAGNGKPLKGPDIYHOTWARYIVKFLDAVAEHK
 LQWVNAENPESAGLSGYPFQCLGTPHQDFIARDLGPLANGTHNVRLLMLD
 DORLLHPWAKVIVLDEPAKIVHGIYVHTLDEPAKATLRETHLFPNTMLFASE
 ACVSGKFWQSVRLSGMDRGMQYSHSITNLLYHVGVGTWNPSL"
 polyA_site 1823
 BASE COUNT 435 a 525 c 489 g 374 t
 ORIGIN
 Query Match 67.8%; Score 1544.2; DB 9; Length 1823;
 Best Local Similarity 95.4%; Pred. No. 0;
 Matches 1623; Conservative 0; Mismatches 23; Indels 55; Gaps 1;
 QY 558 AGGAATCGGATATACATCATCCGGTACCATGCCAGCTGTGACTTCTCCATCCGAC 617
 |||||||
 Db 178 AGGAATCGGATATACATCATCCGGTACCATGCCAGCTGTGACTTCTCCATCCGAC 237
 |||||||
 QY 618 CTACACCTATGCAGACACCCCTGTATTCAGTTGACAACTTCACGCCCTCCAGAGGA 677
 |||||||
 Db 238 CTACACCTATGCAGACACCCCTGTATTCAGTTGACAACTTCACGCCCTCCAGAGGA 297
 |||||||
 QY 678 AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCGATGGCCCGAGCGTCCCGT 737
 |||||||

Db 298 AGATACCAAGCTCAAGATACCCCTGATTCACCGAGCCCTGCAGTTGGCCCGAGCGTCCCGT 357
 |||||||
 QY 738 TTCACTCTCTTCCAGCCCTGGACATCACCCACTTGGCTCAAGACCAATAGGACGGTGAA 797
 |||||||
 Db 358 TTCACTCTCTTCCAGCCCTGGACATCACCCACTCGGCTCAAGACCGAGGGGAGCGGGAA 417
 |||||||
 QY 798 TGGGAAGGGTCACTCAAGGGAGAGCCCGGAGACATCTACCACAGACCTGCGGCCAGATA 857
 |||||||
 Db 418 TGGGAAGGGGCACTCAAGGGAGAGCCCGGAGACATCTACCACAGACCTTGGGCCAGATA 477
 |||||||
 QY 858 CTTTGTGAAGTTCTGTGATGCCTTGTGAGACCAAGTTACAGTTCTGGGCACTGACAGC 917
 |||||||
 Db 478 CATTTGTAAGTTCTGTGATGCCTTGTGAGACCAAGTTACAGTTCTGGGCACTGACAGC 537
 |||||||
 QY 918 TGAATGAGCCTTCTGCTGGGCTTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 977
 |||||||
 Db 538 TGAATGAGCCTTCTGCTGGGCTTGTGAGTGGATACCCCTTCCAGTGCCTGGGCTTCAC 597
 |||||||
 QY 978 CCTGAACATCAGCGAGACTTCATTGCCCCTGACCTAGGTCCTTACCTTGGCAACGGTAC 1037
 |||||||
 Db 598 CCTGAACATCAGCGAGACTTCATTGCCCCTGACCTAGGTCCTTACCTTGGCAACGGTAC 657
 |||||||
 QY 1038 TCACCACAATGTCCGCTACTCATGCTGGATGATGACCAACGCTTGTGCTGCCCACTGGGC 1097
 |||||||
 Db 658 TCACCACAATGTCCGCTACTCATGCTGGATGATGACCAACGCTTGTGCTGCCCACTGGGC 717
 |||||||
 QY 1098 AAAGTGTGTTACTGACAGACCCAGAACGAGCTTAATATGTTTCATGGGATTCCTGTACATTG 1157
 |||||||
 Db 718 AAAGTGTGTTACTGACAGACCCAGAACGAGCTTAATATGTTTCATGGGATTCCTGTACATTG 777
 |||||||
 QY 1158 GTACCTGAGCTTTTGTGCTCCAGCCAAAGCCACCTTAGGGGAGACACACGCTGTTCCT 1217
 |||||||
 Db 778 GTACCTGAGCTTTTGTGCTCCAGCCAAAGCCACCTTAGGGGAGACACACGCTGTTCCT 837
 |||||||
 QY 1218 CAACACCATGCTCTTTCCTCAGAGGCTGTGTTGGGCTCCAAAGTTCTGGGAGCAGAGTGT 1277
 |||||||
 Db 838 CAACACCATGCTCTTTCCTCAGAGGCTGTGTTGGGTTCCAAAGTTCTGGGAGCAGAGTGT 897
 |||||||
 QY 1278 GCGGTAGGCTCCTTGGATCGAGGATGAGTACAGCCACAGCATCATACAGAACCTCCT 1337
 |||||||
 Db 898 GCGGTAGGCTCCTTGGATCGAGGATGAGTACAGCCACAGCATCATACAGAACCTCCT 957
 |||||||
 QY 1338 GTACCATGTGTTGGCTGGACCGACTTGGAACTTTGCCCCGGAAGGAGGAGGCCAA 1397
 |||||||
 Db 958 GTACCATGTGTTGGCTGGACCGACTTGGAA----- 987
 |||||||
 QY 1398 TTGGGTGCGTAACTTTGTCAGCAGTCCCATCTAGACATCACCAGAGCAGCTTTTA 1457
 |||||||
 Db 988 -----CCCATCAATTGTAGACATCACCAGAGCAGCTTTTA 1022
 |||||||
 QY 1458 CAAACAGCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCCA 1517
 |||||||
 Db 1023 CAAACAGCCCATGTTCTACCACTTGGCCACTTCAGCAAGTTTCATTCCTGAGGGCTCCCA 1082
 |||||||
 QY 1518 GAGAGTGGGCTGGTTCGCCAGTCAAGAAAGACGACCTGGAGCGCAGTGGCAGTGTGATGATCC 1577
 |||||||
 Db 1083 GAGAGTGGGCTGGTTCGCCAGTCAAGAAAGACGACCCCGAGCGCAGTGGCAGTGTGATGATCC 1142
 |||||||
 QY 1578 CGATGGCTCTGCTGTTGCTGCTGCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1637
 |||||||
 Db 1143 CGATGGCTCTGCTGTTGCTGCTGCTAAACCGCTCCTCTAAGGATGTGCTCTTACCAT 1202
 |||||||
 QY 1638 CAAGGATCTCTGCTGGGCTTCTGGAGACAATCTCACCTGGCTACTTCCATTTCACACCTA 1697
 |||||||
 Db 1203 CAAGGATCTCTGCTGGGCTTCTGGAGACAATCTCACCTGGCTACTTCCATTTCACACCTA 1262
 |||||||
 QY 1698 CCTGTGCGCTGCGCAGTGTGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGGGCA 1757
 |||||||
 Db 1263 CCTGTGCGCTGCGCAGTGTGAGCAGATACTCAAGGAGGCACTGGGCTCAGCCTGGGCA 1322
 |||||||
 QY 1758 TTAAGGAGCAGAGTCAAGTCAACGCTGTCTGTGACTAAAGAGGCGACAGAGGCGCAG 1817
 |||||||
 Db 1323 TTAAGGAGCAGAGTCAAGTCAACGCTGTCTGTGACTAAAGAGGCGACACAGAGGCGCAG 1382
 |||||||

|||||
Db 1024 CAGCCAAAGCCACCCCTAGGGAGACACACCCTGTTCCCAACACCATGCTCTTTGCC 1083
QY 1237 TCAGAGCCCTGTGTGGCTCCCAAGTTCTGGGACAGAGTGTGGGCTAGGCTCTCTGGGAT 1296
Db 1084 TCAGAGCCCTGTGTGGCTCCCAAGTTCTGGGACAGAGTGTGGGCTAGGCTCTCTGGGAT 1143
QY 1297 CGAGGATCGAGTACAGCCACAGCAGATCATCAGAACCTCTCTGATACATGTGTGGCTGG 1356
Db 1144 CGAGGATCGAGTACAGCCACAGCAGATCATCAGAACCTCTCTGATACATGTGTGGCTGG 1203
QY 1357 ACCGACTGAACCTTGCCCTGAACCCGGAAGGAGGAGGCCAATTGGTGGTAACTTTGTC 1416
Db 1204 ACCGACTGAACCTTGCCCTGAACCCGGAAGGAGGAGGCCAATTGGTGGTAACTTTGTC 1263
QY 1417 GACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTTTACAAACAGCCCATGTTCTAC 1476
Db 1264 GACAGTCCCATCATTTGTAGACATCACCAAGGACAGCTTTTACAAACAGCCCATGTTCTAC 1323
QY 1477 CACCTTGGCCATTCAGCAAGTTTCATCTGAGGGCTCCAGAGAGTGGGGTGGTGGC 1536
Db 1324 CACCTTGGCCATTCAGCAAGTTTCATCTGAGGGCTCCAGAGAGTGGGGTGGTGGC 1383
QY 1537 AGTCAGAAAGACCTGGACGACGTGGCAGTGCATGATGCCGATGCCGCTGCTGCTGTG 1596
Db 1384 AGTCAGAAAGACCTGGACGACGTGGCAGTGCATGATGCCGATGCCGCTGCTGCTGTG 1443
QY 1597 GTCTGCTAAACCGCTCTCTAAGGATGTGCTTACCATCAAGATCCTGCTGTGGGC 1656
Db 1444 GTCTGCTAAACCGCTCTCTAAGGATGTGCTTACCATCAAGATCCTGCTGTGGGC 1503
QY 1657 TTCTGGAGACAATCTACCTGGCTACTCCATTCACACTACCTGTGGCTGGCCAGTGA 1716
Db 1504 TTCTGGAGACAATCTACCTGGCTACTCCATTCACACTACCTGTGGCTGGCCAGTGA 1563

RESULT 11
AX191972
LOCUS AX191972 1551 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 2 from Patent WO0149830.
ACCESSION AX191972
VERSION AX191972.1 GI:15210120
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Okkels, J.S., Jensen, A.D., Halkier, T., Jensen, R.B. and Schambye, H.T.
TITLE Improved lysosomal enzymes and lysosomal enzyme activators
JOURNAL Patent: WO 0149830-A 2 12-JUL-2001;
Maxygen Aps (DK)
FEATURES
source
1..1551
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 345 a 470 c 389 g 347 t
ORIGIN

Query Match 67.7%; Score 1543; DB 6; Length 1551;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1546; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 166 ATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 225
Db 1 ATGGCTGGCAGCCTCACAGGATGCTTCTACTTCAGGCAGTGTGCTGGGCATCAGGTGCC 60
QY 226 GCGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGCTGCAATGCCACA 285
Db 61 GCGCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGTGTGCTGCAATGCCACA 120
QY 286 TACTGTGACTCTCTTGACCCCGACCTTTCTGCTGCTTGTGCTTACCTTCAGCCGCTATGAG 345
|||||

Db 121 TACTGTGACTCTCTTTGACCCCGACACCTTCTCTGCCCTTGGTACCTTCACGCCGCTATGAG 180
QY 346 AGTACACGAGTGGGCGACGAGTGGAGCTGAGTATGGGGCCCATCCAGGCTAAATCAACAG 405
Db 181 AGTACACGAGTGGGCGACGAGTGGAGCTGAGTATGGGGCCCATCCAGGCTAAATCAACAG 240
QY 406 GGCACAGCCCTGCTACTGACCCCTGCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTT 465
Db 241 GGCACAGCCCTGCTACTGACCCCTGCAGCCAGACAGAAAGTTCAGAAAGTGAAGGGATTT 300
QY 466 GGAGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCTGTCAACCCCTGCCCAA 525
Db 301 GGAGGGCCATGACAGATGCTGTGCTCTCAACATCCTTGCCTGTCAACCCCTGCCCAA 360
QY 526 AATTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCCGGTA 585
Db 361 AATTGCTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATACATCATCCCGGTA 420
QY 586 CCATGGCCAGCTGTGACTTCTCCATCCGCACTTACACCTATGACAGACACCCCTGATGAT 645
Db 421 CCCATGGCCAGCTGTGACTTCTCCATCCGCACTTACACCTATGACAGACACCCCTGATGAT 480
QY 646 TTCCAGTTGACAACTTTCAGCCTCCAGAGAGAGATACCAAGCTCAAGATACCCCTGATTT 705
Db 481 TTCCAGTTGACAACTTTCAGCCTCCAGAGAGAGATACCAAGCTCAAGATACCCCTGATTT 540
QY 706 CACGAGCCCTGCAAGTTGGCCCGACGCTCCGCTTCTACTCTTCCAGCCCTGGACATCA 765
Db 541 CACGAGCACTGCAAGTTGGCCCGACGCTCCGCTTCTACTCTTCCAGCCCTGGACATCA 600
QY 766 CCCACTTGGCTCAAGCAATGAGCGGTGAATGGGAAGGGGTCACTCAAGGAGACGCC 825
Db 601 CCCACTTGGCTCAAGCAATGAGCGGTGAATGGGAAGGGGTCACTCAAGGAGACGCC 660
QY 826 GGAGACATCTACCAACAGACCTGSGCCAGATCTTTCTGGAAGTTCCTGGATGCTATGCT 885
Db 661 GGAGACATCTACCAACAGACCTGSGCCAGATCTTTCTGGAAGTTCCTGGATGCTATGCT 720
QY 886 GAGCACAAGTTACAGTTTCTGGGAGTGCAGAGCTGACAGCTGAAATGAGCTTCTGCTGGCTGTG 945
Db 721 GAGCACAAGTTACAGTTTCTGGGAGTGCAGAGCTGACAGCTGAAATGAGCTTCTGCTGGCTGTG 780
QY 946 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAACATCAGGAGACTTCATTTGCC 1005
Db 781 AGTGGATACCCCTTCCAGTGCCTGGGCTTCAACCCCTGAACATCAGGAGACTTCATTTGCC 840
QY 1006 CGTGACCTAGGTCTCTACCCCTGCGCAACAGTACTCACACAATGTCGCCCTTACTCATGCTG 1065
Db 841 CGTGACCTAGGTCTCTACCCCTGCGCAACAGTACTCACACAATGTCGCCCTTACTCATGCTG 900
QY 1066 GATGACCAACGCTTGTGCTGCTGCCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAAGCA 1125
Db 901 GATGACCAACGCTTGTGCTGCTGCCCCCACTGGGCAAGGTGGTACTGACAGACCCAGAAGCA 960
QY 1126 GCTAAATATGTTATGCGCATTTGCTGATGTTGTTGCTGAGCTTCTGCTCCAGCCAA 1185
Db 961 GCTAAATATGTTATGCGCATTTGCTGATGTTGTTGCTGAGCTTCTGCTCCAGCCAA 1020
QY 1186 GGCACCCCTAGGGAGACACACCCGCTGTTCCCAACACCATGCTTCTGCTCCAGAGCC 1245
Db 1021 GGCACCCCTAGGGAGACACACCCGCTGTTCCCAACACCATGCTTCTGCTCCAGAGCC 1080
QY 1246 TGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCTGGATCGAGGGATG 1305
Db 1081 TGTGTGGGCTCCAAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCTGGATCGAGGGATG 1140
QY 1306 CAGTACAGCCACAGCATCATCAGAACCTCTCTGTACCATGTGCTGGCTGGACCGACTGG 1365
Db 1141 CAGTACAGCCACAGCATCATCAGAACCTCTCTGTACCATGTGCTGGCTGGACCGACTGG 1200
QY 1366 AACTTCCCTGAAACCCCGAAGGAGGAGCCCAATTGGGTGGCTAACTTTGTTCGACAGTCC 1425
Db 1201 AACTTCCCTGAAACCCCGAAGGAGGAGCCCAATTGGGTGGCTAACTTTGTTCGACAGTCC 1260
|||||

QY	1426	ATCATTTGTAGACATACCAAGGACACAGT	TTTTACAAACAGAGCCATGTTCTTACCAACCTTGGC	1485
DB	1361	ATCATTTGTAGACATACCAAGGACAGGTTT	TACAAACAGAGCCATGTTCTTACCAACCTTGGC	1320
QY	1486	CACCTTCAGCAAGTTCATTCTCGAGGGCT	CCCCAGAGAGTGGGCTGGTCCGACAGCAAG	1545
DB	1321	CATTTCAGCAAGTTCATTCTCGAGGGCT	CCCCAGAGAGTGGGCTGGTCCGACAGCAAG	1380
QY	1546	AACGACCTGGAGCGAGTGGCACTGATGCAT	CCGAGTGGCTCTGCTGTGTGTGCTGCTA	1605
DB	1381	AACGACCTGGAGCGAGTGGCACTGATGCAT	CCGAGTGGCTCTGCTGTGTGTGCTGCTA	1440
QY	1606	AACGGCTCTCTAAGAGTGTGCTCTTACCAT	CACAGATTCCTGCTGTGGGCTTCCTGGAG	1665
DB	1441	AACGGCTCTCTAAGAGTGTGCTCTTACCAT	CACAGATTCCTGCTGTGGGCTTCCTGGAG	1500
QY	1666	ACAATCTCACCTGGCTACTCCATTCACACT	CACTGTGGCGTCGCCAGTGA 1716	
DB	1501	ACAATCTCACCTGGCTACTCCATTCACACT	TACCTGTGGCGTCGCCAGTGA 1551	
RESULT 12				
LOCUS	AX349364	1551 bp	DNA	PAT 06-FEB-2002
DEFINITION	Sequence 2 from Patent WO0202597.			
ACCESSION	AX349364			
VERSION	AX349364.1	GI:18615327		
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	1			
TITLE	Okkels, J. S., Jensen, A. D. and van den Hazel, B. C.			
JOURNAL	peptide extended glycosylated polypeptides			
FEATURES	Patent: WO 0202597-A, 2 10-JAN-2002;			
source	Maxygen Aps (DK) ; Maxygen Holdings Ltd (KY)			
	Location/Qualifiers			
	1 1551			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
BASE COUNT	345 a 470 c 389 g 347 t			
ORIGIN				
Query Match	67.78; Score 1543; DB 6; Length 1551;			
Best Local Similarity	99.78; Pred. No. 0;			
Matches 1546; Conservative	0; Mismatches 5; Indels 0; Gaps 0;			
QY	166	ATGGCTGGAGCGCTACAGAGTTCCTTCTACT	TCAGGCAGTCTCTGGGCATCAGGTGCC	225
DB	1	ATGGCTGGAGCGCTACAGAGTTCCTTCTACT	TCAGGCAGTCTCTGGGCATCAGGTGCC	60
QY	226	CGCCCTCGATCCCTAAAGCTTCGGCTACAG	CTCGGTGGTGTGCTGCAATGCCACA	285
DB	61	CGCCCTCGATCCCTAAAGCTTCGGCTACAG	CTCGGTGGTGTGCTGCAATGCCACA	120
QY	286	TACTGTGACTCTTTTGACCCCGGACCTTT	TCCTTGGCTTTGGTACCTTCAGCGCTATGAG	345
DB	121	TACTGTGACTCTTTTGACCCCGGACCTTT	TCCTTGGCTTTGGTACCTTCAGCGCTATGAG	180
QY	346	AGTACACGAGTGGCGACGATGGAGCTGAGT	ATATGGGCCCATCCAGGCTAAATCACAG	405
DB	181	AGTACACGAGTGGCGACGATGGAGCTGAGT	ATATGGGCCCATCCAGGCTAAATCACAG	240
QY	406	GGCAGAGCTTCACCTGACCCGACGACGAC	AGATTCACAGAAAGTGAAGGATTT	465
DB	241	GGCAGAGCTTCACCTGACCCGACGACGAC	AGATTCACAGAAAGTGAAGGATTT	300
QY	466	GGAGGGCCATGACAGATGTGCTGCTCTCA	CATCTTCGCCCTGTACCCCTTGCCCCAA	525
DB	301	GGAGGGCCATGACAGATGTGCTGCTCTCA	CATCTTCGCCCTGTACCCCTTGCCCCAA	360

|||||
Db 1441 AACCGCTCTCTAAGGATGCTCTTACCATCAAGGATCCCTGCTGGCTTCTCGGAG 1500
QY 1666 ACAATCTACCTGGCTGCTTCAATTCACACCTACCTGTCGCGCCAGTGA 1716
Db 1501 ACAATCTACCTGGCTGCTTCAATTCACACCTACCTGTCGCGCCAGTGA 1551
RESULT 13
LOCUS AX299777 1592 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 1 from Patent WO0177307.
ACCESSION AX299777
VERSION AX299777.1 GI:17129298
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE Berent, S.L.
AUTHORS Expression system for efficiently producing clinically effective
TITLE lysosomal enzymes (glucocerebrosidase)
JOURNAL Patent: WO 0177307-A 1 18-Oct-2001;
CYTOTOXIC PHARMACEUTICS, INC. (US)
FEATURES
source Location/Qualifiers
1..1592
/organism="Homo sapiens"
/db_xref="taxon:9606"
13..1563
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD12720.1"
/db_xref="GI:17129299"
translation="MAGSLTGLLLQAVSWASGARPQIPKSGYSSVVVCNATYCD
SDPFPALGTFSEYSESRSMELSMGPQANHTGTGLLTLQPEKFKVKGFG
AMTDAALNIALSPAPNLLKSTFSEEGIGNIRVPMASCDPSIRTYADI PDD
FQHNFSLEPDEKILPIHRAQLAQPVSLLASPTPTWLKNGAVNGKSLKG
QGDYTHQWARYFKFLDAYAEHLQFWAYTAEENPSAGLLSGYPCQCLGTFEQR
DFIARDLGPTTGLNTHNVRILMLDQKLLPHWAKVRLDPEAAKYVHGLAVHWYLD
FLPAKATLGTHRFIPNTMLFASACVSKFWEOSVRLGSDRMQVSHSIIINLLY
HVGWTDNIALNPSSGPNWVRNFVDSPIIVDITKDFYKOPMEVHLGHFSEFPEGS
QKVLVSAQKNDLDAVALMHPDGSVAVVVLNRSKSDVPLTIKDPANVLELISPGYSI
HYLWRRQ"
BASE COUNT 359 a 481 c 397 g 355 t
ORIGIN
Query Match 67.1%; Score 1529.6; DB 6; Length 1592;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1541; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 157 GTAAGCATCATGCTGGCAGCCTCACAGGATTGCTTCTACTTCAGGCAGTGTGCTGGCA 216
Db 4 GAATTCANAATGCTGGCTGCTTAACAGGATTGCTTCTACTTCAGGCAGTGTGCTGGCA 63
QY 217 TCAGGTGCGCGCCCGCTGCTATCCCTAAAGCTTCGGGTACAGCTCGGTGTGTCTGTC 276
Db 64 TCCGGAGCTAGACCTTGCATCCCTTAAAGCTTCGGGTACAGCTCGGTGTGTGTCTGTC 123
QY 277 AATGCCACATACGTGACCTCTTTGACCCCGCCGACCTTTCCTGCGCCCTTGGTACCTTCAGC 336
Db 124 AATGCCACATACGTGACCTCTTTGACCCCGCCGACCTTTCCTGCGCCCTTGGTACCTTCAGC 183
QY 337 CGCTATCAGAGTACAGCAGTGGCGGACGAGTGGAGCTGAGTATGGGGCCCATCCAGGCT 396
Db 184 CGCTATGAGAGTACAGCAGTGGCGGACGAGTGGAGCTGAGTATGGGGCCCATCCAGGCT 243
QY 397 AATCACAGGGGACAGGCTGCTGCTGACCTGCGACGCCAGACAGAGTTCAGAAAGTG 456
Db 244 AATCACAGGGGACAGGCTGCTGCTGACCTGCGACGCCAGACAGAGTTCAGAAAGTG 303
QY 457 AAGGGATTTGGAGGGCCATGACAGATGCTGCTGCTCAACATCTTCCCTGTCACACC 516
|||||

Db 304 AAGGATTTGGAGGGGCCATGACAGATGCTGCTCTCAACATCCCTGCGCTGTACCC 363
QY 517 CCTCCCAAAATTTGCTACTTAAATCGTACTTCTCTCAAGAAGGAATCGGATATAACATC 576
Db 364 CCTCCCAAAATTTGCTACTTAAATCGTACTTCTCTCAAGAAGGAATCGGATATAACATC 423
QY 577 ATCCGGGTACCATGCGGACGTGTGACTTCTCCATCCGACCTACACCTATGAGACACC 636
Db 424 ATCCGGGTACCATGCGGACGTGTGACTTCTCCATCCGACCTACACCTATGAGACACC 483
QY 637 CCTCATGATTTCCAGTTGCGACAACCTTCAGCTCCAGGAAAGATACCAAGCTCAAGATA 696
Db 484 CTTGATGATTTCCAGTTGCGACAACCTTCAGCTCCAGGAAAGATACCAAGCTCAAGATA 543
QY 697 CCCCTGATTTACCAGGACCTTCAGTTGGCCAGCGTCCCGTTTTCACCTTCTGCGAGCCC 756
Db 544 CCCCTGATTTACCAGGACCTTCAGTTGGCCAGCGTCCCGTTTTCACCTTCTGCGAGCCC 603
QY 757 TGGACATACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGGTCTACTCAAG 816
Db 604 TGGACATACCCACTTGGCTCAAGACCAATGGAGCGGTGAATGGGAGGGGTCTACTCAAG 663
QY 817 GGACAGCCCGGAGACATCTACCCAGACCTGGGCCAGATCTTTGTGAAGTTCTCTGAT 876
Db 664 GGACAGCCCGGAGACATCTACCCAGACCTGGGCCAGATCTTTGTGAAGTTCTCTGAT 723
QY 877 GCCTATGCTGAGCACAAAGTTACAGTTCTGGGACGTGACAGCTGAAATGAGCCCTTCTGCT 936
Db 724 GCCTATGCTGAGCACAAAGTTACAGTTCTGGGACGTGACAGCTGAAATGAGCCCTTCTGCT 783
QY 937 GGGCTGTTGAGTGATACCCCTTCAGTGGCTGGCTTACCCCTGACACATCAGCGAGAC 996
Db 784 GGGCTGTTGAGTGATACCCCTTCAGTGGCTGGCTTACCCCTGACACATCAGCGAGAC 843
QY 997 TTCTATGCTGACCTAGGTGCTTACCTCCCAACAGTACTCACCACCAATGTCCGCTA 1056
Db 844 TTCTATGCTGACCTAGGTGCTTACCTCCCAACAGTACTCACCACCAATGTCCGCTA 903
QY 1057 CTCTATGCTGATGACCAACAGCTTCTGCTGCTGCGCCCAAGAGTGGTACTGACAGAC 1116
Db 904 CTCTATGCTGATGACCAACAGCTTCTGCTGCTGCGCCCAAGAGTGGTACTGACAGAC 963
QY 1117 CCAGACGAGCTAATATGTTGATGGCATGCTGTACATGTTGGTACCTGGACTTCTGCTGCT 1176
Db 964 CCAGACGAGCTAATATGTTGATGGCATGCTGTACATGTTGGTACCTGGACTTCTGCTGCT 1023
QY 1177 CCAGCAAGGACACCTAGGGGAGACACACGCTGTTTCCCAACACATGCTCTTTGCC 1236
Db 1024 CCAGCAAGGACACCTAGGGGAGACACACGCTGTTTCCCAACACATGCTCTTTGCC 1083
QY 1237 TCAGAGGCTGTGTGGGCTCCAAAGTTCTGGGAGAGAGTGTGCGGCTAGGCTCCTGGAT 1296
Db 1084 TCAGAGGCTGTGTGGGCTCCAAAGTTCTGGGAGAGAGTGTGCGGCTAGGCTCCTGGAT 1143
QY 1297 CGAGGATGAGTACAGCCACAGCATATCAGAACCTTCTGTACCATGTTGGTGGCTGG 1356
Db 1144 CGAGGATGAGTACAGCCACAGCATATCAGAACCTTCTGTACCATGTTGGTGGCTGG 1203
QY 1357 ACCGACTGGAACCTTGGCCCTGAAACCCCGGAGGAGGACCAATTTGGTGGCTTTGTC 1416
Db 1204 ACCGACTGGAACCTTGGCCCTGAAACCCCGGAGGAGGACCAATTTGGTGGCTTTGTC 1263
QY 1417 GACAGTCCCATCATTTGATAGATCACCAGACAGCTTTTACAAACAGCCCATGTTCTAC 1476
Db 1264 GACAGTCCCATCATTTGATAGATCACCAGACAGCTTTTACAAACAGCCCATGTTCTAC 1323
QY 1477 CACCTTGCCACTTTCAGCAAGTTTATCTCCAGGCTCCAGAGAGTGGGCTGGTGGC 1536
Db 1324 CACCTTGCCACTTTCAGCAAGTTTATCTCCAGGCTCCAGAGAGTGGGCTGGTGGC 1383
QY 1537 AGTCAGAAAGACGACCTGGACGAGTGGCACTGATGATCCCGATGCGTCTGCTGTGTG 1596
Db 1384 AGTCAGAAAGACGACCTGGACGAGTGGCACTGATGATCCCGATGCGTCTGCTGTGTG 1443

QY 1597 GTCGTGCTAAACCGCTCTCTAAGGATGCGCTCTTACCATCAAGGATCCTGCTGTGGC 1656
 Db 1444 GTCGTGCTAAACCGCTCTCTAAGGATGCGCTCTTACCATCAAGGATCCTGCTGTGGC 1503
 QY 1657 TTCCTGAGACAAATTCACCTGCTACTCTCAATTCACACCTACCTGCGCGCTGAGTGA 1716
 Db 1504 TTCCTGAGACAAATTCACCTGCTACTCTCAATTCACACCTACCTGCGCGCTGAGTGA 1563

RESULT 14
 BC000349
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BC000349 1900 bp mRNA linear PRI 12-JUL-2001
 Homo sapiens, clone MGC:8614 IMAGE:2961326, mRNA, complete cds.
 BC000349.1 GI:12653162
 MGC.
 Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1900)
 Strausberg, R.
 Direct Submission
 Submitted (15-NOV-2000) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LINL)
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stantripo, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRAL Plate: 1 Row: m Column: 5
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 496370.
 Location/Qualifiers
 1. 1900
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:8614 IMAGE:2961326"
 /tissue_type="Muscle, rhabdomyosarcoma"
 /clone_lib="NIH_MGC_17"
 /lab_host="DH10B-R"
 /note="Vector: pOT7"
 377..1339
 /codon_start=1
 /product="Unknown (protein for MGC:8614)"
 /protein_id="AAH00349.1"
 /db_xref="gi:12653162"
 /translation="MASCDFSTRTYTYADTPDFQLHNFSLPEEDTKLPLIHLALQ
 LAQPVSLIAPSPTRKLTGKAGKGLKQPRDIYHQTWARYIVKFLDAYAEHK
 LQFVATNENPFSAGLSYFQCLGTFEHOEDFIARDLGTLANGTHTHHPVPCGR
 DQRLPLIPAKVVLTPDEAAKGLCGVFLGAEARLSDVQPDQHHKHPVPCGR
 LDRLPEIFVDITKTFQKPMFVHLGHSKFIPEGSQVGLVAVSQNDPAVALMHPD
 GSPVVVLNRSKSDVPITIKDPAVGLPFIETISPGYSIHTLWRRQ"

471 a 524 c 506 g 399 t

ORIGIN
 Query Match 58.9%; Score 1343; DB 9; Length 1900;
 Best Local Similarity 59.0%; Pred. No. 0;
 Matches 1943; Conservative 0; Mismatches 25; Indels 165; Gaps 2;

QY 545 ACTTCTCTGAAGAAGGAATCGGATATACATCATCCGGGTACCATGCCAGCTGTGACT 604
 Db 333 ATTGCCCGGACAGGAATCGGATATACATCATCTGGTACCATGCCAGCTGTGACT 392
 QY 605 TCTCATCCGACCTACATCTATGACAGACCCCTGATGATTTCCAGTTGCAACACTTCA 664
 Db 393 TCTCATCCGACCTACATCTATGACAGACCCCTGATGATTTCCAGTTGCAACACTTCA 452
 QY 665 GCCTCCGACGGAAGATCAAGCTCAAGATACCTGATTCACCGAGCCCTGAGTTGG 724
 Db 453 GCCTCCGACGGAAGATCAAGCTCAAGTACCTGATTCACCGAGCCCTGAGTTGG 512
 QY 725 CCGACGTCGCGTTTCACTCTCTGCGAGCCCTGACATCACCCTGCTGCTCAAGACCA 784
 Db 513 CCGACGTCGCGTTTCACTCTCTGCGAGCCCTGACATCACCCTGCTGCTCAAGACCA 572
 QY 785 ATGGACGCTGAATGGGAAGGCTGACTCAAGGACAGCCGAGACATACACACAGA 844
 Db 573 AGGGAGCGGGAATGGGAAGGCTGACTCAAGGACAGCCGAGACATACACACAGA 632
 QY 845 CCGTGGCCAGATCTTTGTAAGTTCTCTGATGCTGATGCTGATGCTGATGCTGATGCT 904
 Db 633 CCGTGGCCAGATCTTTGTAAGTTCTCTGATGCTGATGCTGATGCTGATGCTGATGCT 592
 QY 905 GGGCAGTGACAGCTGAAATGAGCCCTCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCT 964
 Db 693 GGGCAGTGACAGCTGAAATGAGCCCTCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCT 752
 QY 965 GCGTGGGCTTCAACCTGACATCAGCAGACTTCAATGCCCCGCTGCTGCTGCTGCTGCT 1024
 Db 753 GCGTGGGCTTCAACCTGACATCAGCAGACTTCAATGCCCCGCTGCTGCTGCTGCTGCT 812
 QY 1025 TCGCCACAGTACTACCAATGTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1084
 Db 813 TTGCCAAGGCTACTACCAATGTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 872
 QY 1085 TGCCCCACTGGGCAAGGTTGGTACTGACAGACCCAGAGAGCAATATGTTATGTCATGGCA 1144
 Db 873 TGCCCCACTGGGCAAGGTTGGTACTGACAGACCCAGAGAGCAATATGTTATGTCATGGCA 918
 QY 1145 TTGCTGTACATTTGTTACTTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
 Db 919 -----AGGCCCTGTGTGGTTCCTCAAGTCT 1264
 QY 1205 ACCGCTGTTCCTCAAGACCATGCTCTTTGCTCAGAGGCTGTGTGGGCTCCCAAGTTCT 1264
 Db 919 -----AGGCCCTGTGTGGTTCCTCAAGTCT 942
 QY 1265 GGGAGCAGAGTGTGCGCTAGGCTCCTGGGATCGAGAGGATCGAGTACAGCAGCAGCATCA 1324
 Db 943 GGGAGCAGAGTGTGCGCTAGGCTCCTGGGATCGAGAGGATCGAGTACAGCAGCAGCATCA 1002
 QY 1325 TCAGAACCTCCTGTTACCATGTGGTGGGCTGGAGCTTGAACCTTGCCTTGAACCCGG 1384
 Db 1003 TCAGAACCTCCTGTTACCATGTGGTGGGCTGGAGCTTGAACCTTGCCTTGAACCCGG 1045
 QY 1385 AAGGAGGACCCCAATTGGGTGCTTAATTTGTCGAGTCCCATCATTTGATAGACATCACCA 1444
 Db 1046 -----CCCATCATTTGATAGACATCACCA 1067
 QY 1445 AGGACAGTTTACAAACACCCCATGTTCTACACCTTGGCCACTTGCAGCAAGTTCAATTC 1504
 Db 1068 AGCAGAGTTTACAAACACCCCATGTTCTACACCTTGGCCACTTGCAGCAAGTTCAATTC 1127
 QY 1505 CTGAGGCTCCCAAGAGAGTGGGCTGGTGGCAGTCAAGAGACGACCTGGAGCGCAGTGG 1564
 Db 1128 CTGAGGCTCCCAAGAGAGTGGGCTGGTGGCAGTCAAGAGACGACCGCGCAGTGG 1187

FEATURES
 source

CDS

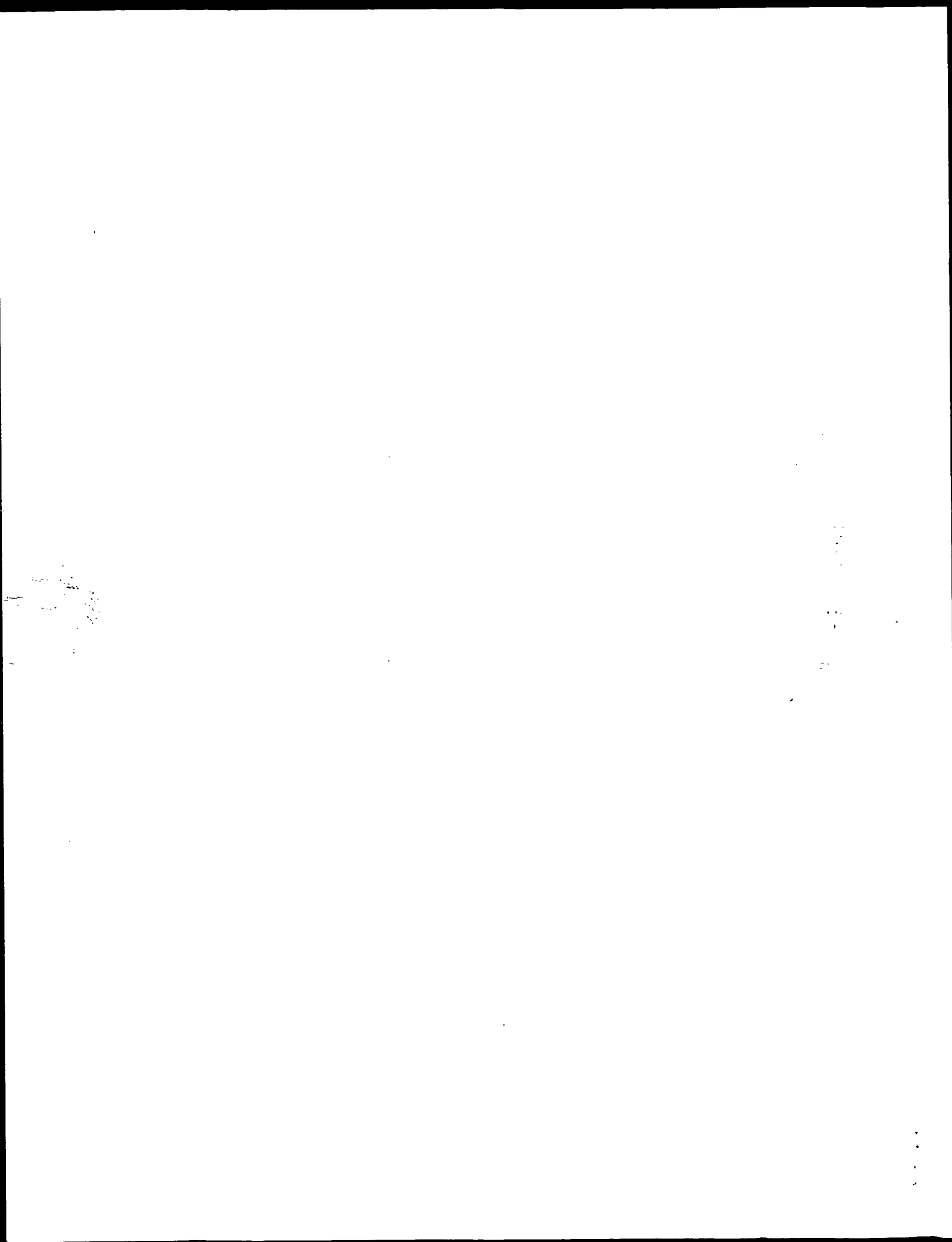
BASE COUNT

QY 1565 CACTGATGATCCCGATGCTCTGCTGTGTGTGCTGCTTAACCGCTCCTCTTAAGGATG 1624
 Db 1188 CACTGATGATCCCGATGCTCTGCTGTGTGTGCTGCTTAACCGCTCCTCTTAAGGATG 1247
 QY 1625 TGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTGAGGACATCTCACTGGCTACT 1684
 Db 1248 TGCTCTTACCATCAAGGATCTGCTGTGGGCTTCTGAGGACATCTCACTGGCTACT 1307
 QY 1685 CATTCACACCTACCTGTGGGCTGCCAGTGATGAGCAGATCTCAAGGAGGACCTGGG 1744
 Db 1308 CATTCACACCTACCTGTGGGCTGCCAGTGATGAGCAGATCTCAAGGAGGACCTGGG 1367
 QY 1745 CTCAGCTGGGCAATTAAGGGACAGATCAGCTCAGCTCAGCTGTCTGTGACTTAAGAGGCG 1804
 Db 1368 CTCAGCTGGGCAATTAAGGGACAGATCAGCTCAGCTCAGCTGTCTGTGACTTAAGAGGCG 1427
 QY 1805 ACAGCAGGGCCAGTGTGAGCTTACAGCAGCAGTAAAGCCAGGGGCAATGTTGGTGACT 1864
 Db 1428 ACAGCAGGGCCAGTGTGAGCTTACAGCAGCAGTAAAGCCAGGGGCAATGTTGGTGACT 1487
 QY 1865 CACTTTCCCTCTAGTGTGTGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1924
 Db 1488 CACTTTCCCTCTAGTGTGTGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1547
 QY 1925 CAGTGTCCCTCCAGCCCATCTTATGTGAATGCGCTGTGCTGTGCTGTGCTTGTGAA 1984
 Db 1548 CAGTGTCCCTCCAGCCCATCTTATGTGAATGCGCTGTGCTGTGCTGTGCTTGTGAA 1607
 QY 1985 ACTGGGCTGGTCCAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 2044
 Db 1608 ACTGGGCTGGTCCAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 1667
 QY 2045 TGAGGTAAGGAAAGAGAGACTAGGAAAGCTGGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 2104
 Db 1668 TGAGGTAAGGAAAGAGAGACTAGGAAAGCTGGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1727
 QY 2105 TTCTGGAGATGAGACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 2164
 Db 1728 TTCTGGAGATGAGACTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 1787
 QY 2165 TAAAGCAGCAGGCGGTGTGCCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 2224
 Db 1788 TAAAGCAGCAGGCGGTGTGCCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1847
 QY 2225 ATGGCAGCTCTTAAAGGAGAAATGTTGAGCCCAAAAAA 2277
 Db 1848 ATGGCAGCTCTTAAAGGAGAAATGTTGAGCCCAAAAAA 1900
 RESULT 15
 AK090514
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ33195 fis, clone ADRGL2005961, highly similar to GLUCOSYLKERAMIDASE PRECURSOR (EC 3.2.1.45).
 ACCESSION AK090514
 VERSION AK090514.1 GI:21748690
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens adrenal gland cDNA to mRNA, clone_lib:ADRL2
 clone:ADRL2005961.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished

2 (bases 1 to 2711)
 Isogai, T. and Yamamoto, J.
 Direct Submission
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 source
 1..2711
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="ADRL2005961"
 /tissue_type="adrenal gland"
 /clone_lib="ADRL2"
 /notes="cloning vector: pME18SFL3"
 BASE COUNT 622 a 789 c 688 g 612 t
 ORIGIN
 Query Match 56.5%; Score 1287; DB 9; Length 2711;
 Best Local Similarity 85.0%; Pred. No. 0;
 Matches 1517; Conservative 0; Mismatches 195; Indels 73; Gaps 4;
 QY 547 TTCTCTGAAGAAGGAATCGGATATAACATCATCGGTGATCCCATGGCAGCTGTGACTTC 606
 Db 927 TCCCTTTCTGTAGGAATCGGATATAACATCATCTGGGTGATCCCATGGCAGCTGTGACTTC 986
 QY 607 TCCATCCGCACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCAACATTCAGC 666
 Db 987 TCCATCCGCACCTACACCTATGACAGACACCCCTGATGATTTCCAGTTGCAACATTCAGC 1046
 QY 667 CTCCCAGGAGGATACCAAGCTCAAGTACCCCTGATTCACCGAGCCCTGCGAGTTGGCC 726
 Db 1047 CTCCCAGGAGGATACCAAGCTCAAGTACCCCTGATTCACCGAGCCCTGCGAGTTGGCC 1106
 QY 727 CAGCGTCCGCTTTTCTCCTCTGCGAGCCCTTGGACATCATCCCTGCTCAAGACCAAT 786
 Db 1107 CAGCGTCCGCTTTTCTCCTCTGCGAGCCCTTGGACATCATCCCTGCTCAAGACCAAG 1166
 QY 787 GGAGCGGTGAATGGGAAGGGGTCACTCAAGGACAGCCCGGAGACATCTACCACGACC 846
 Db 1167 GGAGCGGTGAATGGGAAGGGGTCACTCAAGGACAGCCCGGAGACATCTACCACGACC 1226
 QY 847 TGGGCCAGATACTTTGTGAAGTTCTCTGATGCCCTATGCTGAGCACAAGTTACAGTTCTGG 906
 Db 1227 TGGGCCAGATACTTTGTGAAGTTCTCTGATGCCCTATGCTGAGCACAAGTTACAGTTCTGG 1286
 QY 907 CAGTGACAGCTGAAAATGAGCCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGC 966
 Db 1287 CAGTGACAGCTGAAAATGAGCCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGC 1346
 QY 967 CTGGGCTTCAACCCCTGAAACATCAGCAGAGCTTCATTCGCCGTGACCTAGTCTACCCCTC 1026
 Db 1347 CTGGGCTTCAACCCCTGAAACATCAGCAGAGCTTCATTCGCCGTGACCTAGTCTACCCCTC 1406
 QY 1027 CCAACAGTACTACCAACAATGTCCGCTACTCATGCTGGATGACCAACGCTTCTGCTGTG 1086
 Db 1407 CCAACAGTACTACCAACAATGTCCGCTACTCATGCTGGATGACCAACGCTTCTGCTGTG 1466
 QY 1087 CCCCACTGGGCAAGGTGGTGTACTGACAGACCCAGAGCAGCTAAATATGTTTCAATGGCATT 1146
 Db 1467 CCCCACTGGGCAAGGTGGTGTACTGACAGACCCAGAGCAGCTAAATATGTTTGGGTTTC 1526
 QY 1147 GCTGTAC-----ATTGGTACCTGGAATTTCTG-----CTCCAGCCA 1183
 Db 1527 CAAGTTCTGGGAGCAGAGTGTGGGCTAGGCTCTGGATCGAGGGATGACGTACAGCCA 1586

QY 1184 AAGCCACCTAGGGAGACACACCGCTCTTCCCAACACCATGCTCTTTGGCTCAGAG 1243
 Db 1587 CAGCATCATCAGCTCTGCACACCTCATTTGCCCTTTTGCACACTACTAGGACGACTTGCAG 1646
 QY 1244 C-----CTGTGTGGGCTCCAGTTCTGGAGCAGAGTGTGGGCTAGGCTCTCTGGGA 1295
 Db 1647 CTGCTCAGACTTCTCAGCTCCCTTGAGATGCTGAGTCTTACACCCCAACTCCTTA 1706
 QY 1296 TCGAGGATGAGTACAGCAGCAGATCATCAGAACCTCTGTACCATGTGTGCGCTG 1355
 Db 1707 GCTACTAAGGAATGTGCCCTCAGGGCTGACCTACCACAGCTGCTCTCCACACGT 1766
 QY 1356 GACCGACTGGAACTTCCCTGTAACCCCGAAGGAGGACCCAAATTGGGTGGCTAACTTTGT 1415
 Db 1767 GACCTTACCTTACACTCTCTGGGGACCCCACTGTGGGCTTGTCTCCTTGCCTTGT 1826
 QY 1416 CGACA-----GTCCCATCATTTGT 1433
 Db 1827 CCTTACCTTAGAACCTCTCTTACCATGTGGTGGACGACTGGAACCCATCATTTGT 1886
 QY 1434 AGCATATCACAAGACACAGCTTTTACAAACAGCCCATTTCTACCACTTGGCCACTTCAG 1493
 Db 1887 AGCATATCACAAGACACAGCTTTTACAAACAGCCCATTTCTACCACTTGGCCACTTCAG 1946
 QY 1494 CAAGTTTCAATTCCTCAGGGCTCCAGAGAGTGGGCTGGTTCGCCAGTCAAGAACGACCT 1553
 Db 1947 CAAGTTTCAATTCCTCAGGGCTCCAGAGAGTGGGCTGGTTCGCCAGTCAAGAACGACCT 2006
 QY 1554 GGACGAGTGGCAGTATGATCCCGATGGCTCTGCTGTGTGTCTGTAAACCGCTC 1613
 Db 2007 GGACGAGTGGCAGTATGATCCCGATGGCTCTGCTGTGTGTCTGTAAACCGCTC 2066
 QY 1614 CTCTAAGATGTGCTCTTACCATCAAGATCTGCTGTGGGCTTCTGGAGACAATCTC 1673
 Db 2067 CTCTAAGATGTGCTCTTACCATCAAGATCTGCTGTGGGCTTCTGGAGACAATCTC 2126
 QY 1674 ACCTGGCTACTTCCATTTACACACTTACCTGTGGCTGGCTGAGTGGAGCAGATACTCAAG 1733
 Db 2127 ACCTGGCTACTTCCATTTACACACTTACCTGTGGCTGGCTGAGTGGAGCAGATACTCAAG 2186
 QY 1734 GAGGCACTGGGCTCAGCTGGGCTTAAAGGACAGAGTCACTCAGCTCAGCTCTCTGTGA 1793
 Db 2187 GAGGCACTGGGCTCAGCTGGGCTTAAAGGACAGAGTCACTCAGCTCAGCTCTCTGTGA 2246
 QY 1794 CTAAGAGGACAGCAGGAGCAGTGTGAGCTTACGCGAGTAAAGCCAGGGCAATGG 1853
 Db 2247 CTAAGAGGACAGCAGGAGCAGTGTGAGCTTACGCGAGTAAAGCCAGGGCAATGG 2306
 QY 1854 TTTGGGTCACTCACTTTCCTCTAGTGTGGTGGCCAGGGCTGGAGGCCCTAGAAAAA 1913
 Db 2307 TTTGGGTCACTCACTTTCCTCTAGTGTGGTGGCCAGGGCTGGAGGCCCTAGAAAAA 2366
 QY 1914 TCAGTAAGCCCAAGTGTCCCCCAAGCCCATGCTTATGTGAACATGCGCTGTGTCTGC 1973
 Db 2367 TCAGTAAGCCCAAGTGTCCCCCAAGCCCATGCTTATGTGAACATGCGCTGTGTCTGC 2426
 QY 1974 TTGCTTTGAAACTGGGCTTGGGCTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACAC 2033
 Db 2427 TTGCTTTGAAACTGGGCTTGGGCTCCAGGCTAGGGTGAAGTCACTGTCCGTACAAACAC 2486
 QY 2034 AAGATCAGGCTCAGGGTAAAGAAAGAGAGACTAGGAAGCTGGGCCCAAACTGGAG 2093
 Db 2487 AAGATCAGGCTCAGGGTAAAGAAAGAGAGACTAGGAAGCTGGGCCCAAACTGGAG 2546
 QY 2094 ACTGTTTGTCTTCCGAGATGAGAACTGGGCCCTGGAGCAGAGTGTGAGCATCAG 2153
 Db 2547 ACTGTTTGTCTTCCGAGATGAGAACTGGGCCCTGGAGCAGAGTGTGAGCATCAG 2606
 QY 2154 GGGGAGAGCTTAAAGCAGCAGCGGTGTGCCAGGCAACCATGATTCCTATGGCACC 2213
 Db 2607 GGGGAGAGCTTAAAGCAGCAGCGGTGTGCCAGGCAACCATGATTCCTATGGCACC 2666
 QY 2214 AGCCAGGAAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGAGCCC 2258

Db -2667 AGCCAGGAAAAATGGCAGCTCTTAAAGGAGAAAAATGTTTGAGCCC 2711
 Search completed: February 22, 2003, 12:19:28
 Job time : 5885 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:17:27 ; Search time 31 Seconds
(without alignments)
508.732 Million cell updates/sec

Title: US-10-024-197-25

Perfect score: 2865

Sequence: 1 MFSPSPRECEPKPLSRVSI.....GFLETISPCYSIHTYLRRO 536

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA:
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2864	100.0	546	2	US-08-713-928B-4		Sequence 4, Appli
2	2865	93.0	498	1	US-08-080-855-1		Sequence 1, Appli
3	105.5	3.7	430	2	US-08-924-440-2		Sequence 2, Appli
4	97	3.4	1711	2	US-08-342-930-2		Sequence 2, Appli
5	94.5	3.3	180	2	US-08-924-759-24		Sequence 24, Appl
6	94.5	3.3	180	3	US-09-248-335-24		Sequence 24, Appl
7	93	3.2	15281	2	US-08-471-119A-2		Sequence 2, Appli
8	92	3.2	716	4	US-08-971-188-5		Sequence 5, Appli
9	92	3.2	716	4	US-08-971-188-11		Sequence 11, Appl
10	90.5	3.2	437	4	US-09-801-052-4		Sequence 4, Appli
11	90.5	3.2	488	4	US-09-801-052-5		Sequence 5, Appli
12	88.5	3.1	4302	3	US-08-658-136-5		Sequence 5, Appli
13	88.5	3.1	4302	4	US-09-052-469-8		Sequence 8, Appli
14	88.5	3.1	4303	2	US-08-460-751-2		Sequence 2, Appli
15	88.5	3.1	4339	4	US-09-052-469-6		Sequence 6, Appli
16	88	3.1	556	2	US-08-016-000-4		Sequence 4, Appli
17	86.5	3.0	2629	2	US-08-751-189-4		Sequence 4, Appli
18	86.5	3.0	2629	2	US-09-060-836-4		Sequence 4, Appli
19	86.5	3.0	2629	4	US-09-184-445-4		Sequence 4, Appli
20	86	3.0	675	4	US-08-485-355B-52		Sequence 52, Appl
21	85.5	3.0	465	4	US-09-801-052-2		Sequence 2, Appli
22	85.5	3.0	591	3	US-08-965-903B-2		Sequence 2, Appli
23	85	3.0	501	3	US-08-906-791-2		Sequence 2, Appli
24	84.5	2.9	728	4	US-08-915-337-2		Sequence 2, Appli
25	84	2.9	545	4	US-09-066-046-39		Sequence 39, Appl
26	83.5	2.9	274	1	US-08-318-947A-20		Sequence 20, Appl
27	83.5	2.9	274	2	US-08-795-303-20		Sequence 20, Appl

28	83	2.9	690	4	US-09-228-986-69	Sequence 69, Appl
29	82	2.9	1323	1	US-08-026-138E-4	Sequence 4, Appli
30	81.5	2.8	737	4	US-08-971-188-4	Sequence 4, Appli
31	81.5	2.8	852	2	US-08-408-519-5	Sequence 5, Appli
32	81.5	2.8	852	5	PCT-US95-03552-5	Sequence 5, Appli
33	81.5	2.8	953	4	US-09-245-281-43	Sequence 43, Appl
34	81.5	2.8	953	4	US-09-207-359B-43	Sequence 43, Appl
35	81	2.8	597	1	US-08-399-646-4	Sequence 4, Appli
36	81	2.8	597	1	US-08-607-321-4	Sequence 4, Appli
37	81	2.8	597	2	US-08-961-240-4	Sequence 4, Appli
38	81	2.8	597	2	US-08-605-501-4	Sequence 4, Appli
39	81	2.8	598	1	US-08-399-646-14	Sequence 14, Appl
40	81	2.8	598	1	US-08-607-321-14	Sequence 14, Appl
41	81	2.8	598	2	US-08-961-240-14	Sequence 14, Appl
42	81	2.8	598	2	US-08-605-501-14	Sequence 14, Appl
43	80.5	2.8	212	2	US-08-924-759-14	Sequence 14, Appl
44	80.5	2.8	212	3	US-09-248-335-14	Sequence 14, Appl
45	80.5	2.8	580	4	US-09-188-930-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-08-713-928B-4
; Sequence 4, Application US/08713928B
; Patent No. 5929304
; GENERAL INFORMATION:
; APPLICANT: RADIN, DAVID N.
; APPLICANT: CRAMER, CAROLE L.
; APPLICANT: OISHI, KAREN K.
; APPLICANT: WEISSBORN, DEBORAH L.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08713,928B
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/003,737
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7956-0011-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-713-928B-4

Query Match 100.0%; Score 2864; DB 2; Length 546;
Best Local Similarity 99.8%; Pred. No. 2.1e-304;

Matches 535; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEFSPPRECKPLSRYSIMAGSLTGLLLQAVSWAGARPCPKPSGYSVVCVNCAT 60
DB 1 MEFSPPRECKPLSRYSIMAGSLTGLLLQAVSWAGARPCPKPSGYSVVCVNCAT 60
QY 61 YCDSFDPPTFPALGTFSTRYSRGRMELSMGPTQANHGTGLLLTLQPEQKQKVGK 120
DB 61 YCDSFDPPTFPALGTFSTRYSRGRMELSMGPTQANHGTGLLLTLQPEQKQKVGK 120
QY 121 GCAMTDAALNIALSPPAQNLLKSYSEEGIGYNIIRVPMASCDFSIRTYTADTDD 180
DB 121 GCAMTDAALNIALSPPAQNLLKSYSEEGIGYNIIRVPMASCDFSIRTYTADTDD 180
QY 181 FOLHNFSLPEEDTKLPIHLIRALQAPVSLASPTWPTWLTNTGAVNGKSLKQOP 240
DB 181 FOLHNFSLPEEDTKLPIHLIRALQAPVSLASPTWPTWLTNTGAVNGKSLKQOP 240
QY 241 GDIYHQTWARYVFKLDAYAEHKLQFWAVTAENEPSAGLLSGYPQCLGFTPEHQDFIA 300
DB 241 GDIYHQTWARYVFKLDAYAEHKLQFWAVTAENEPSAGLLSGYPQCLGFTPEHQDFIA 300
QY 301 RDLGPTLANSTHNNVRLMLDQRLLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
DB 301 RDLGPTLANSTHNNVRLMLDQRLLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
QY 361 ATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
DB 361 ATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
QY 421 NLALNPEGGPNVRFVDSPIIVDTYKTFYKQPMFYHLGHFSKFIPEGSQVGLVASOK 480
DB 421 NLALNPEGGPNVRFVDSPIIVDTYKTFYKQPMFYHLGHFSKFIPEGSQVGLVASOK 480
QY 481 NDLDAVALMHPDGSVVVYVNLNRSSKDVPLTIKDPVAGLETISPYSIHTYLWRRQ 536
DB 481 NDLDAVALMHPDGSVVVYVNLNRSSKDVPLTIKDPVAGLETISPYSIHTYLWRRQ 536

RESULT 2

US-08-080-855-1
; Sequence 1, Application US/08080855
; Patent No. 549892
; GENERAL INFORMATION:
; APPLICANT: FRIEDMAN, BETHANN
; APPLICANT: HAYES, MICHAEL
; TITLE OF INVENTION: ENHANCED IN VIVO UPTAKE OF
; TITLE OF INVENTION: GLUCOCEREEROSIDASE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION / LEGAL DEPARTMENT
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MA
; COUNTRY: USA
; ZIP: 01701-9322
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/080,855
; FILING DATE: 21-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,283
; FILING DATE: 23-DEC-1989
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/455,507
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/289,589

FILING DATE: 23-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: GOSZ ESQ, WILLIAM G
REGISTRATION NUMBER: 27,787
REFERENCE/DOCKET NUMBER: GC22.5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-270-2583
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-080-855-1

Query Match 93.0%; Score 2665; DB 1; Length 498;
Best Local Similarity 99.8%; Pred. No. 1.1e-282;
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 40 ARPCIPKSGYSSVVCVNCATYCDSPDPTFPALGTFSTRYSRGRMELSMGPTQANH 99
DB 1 ARPCIPKSGYSSVVCVNCATYCDSPDPTFPALGTFSTRYSRGRMELSMGPTQANH 60
QY 100 TGTGLLLTLQPEQKQKVGKFGGAMTDAALNIALSPPAQNLLKSYSEEGIGYNIIR 159
DB 61 TGTGLLLTLQPEQKQKVGKFGGAMTDAALNIALSPPAQNLLKSYSEEGIGYNIIR 120
QY 160 VPMASCDFSIRTYTADTDDFOLHNFSLPEEDTKLPIHLIRALQAPVSLASPTW 219
DB 121 VPMASCDFSIRTYTADTDDFOLHNFSLPEEDTKLPIHLIRALQAPVSLASPTW 180
QY 220 SPTWLKTNAGVNGKSLKQOPGDIYHQTWARYVFKLDAYAEHKLQFWAVTAENEPSAGL 279
DB 181 SPTWLKTNAGVNGKSLKQOPGDIYHQTWARYVFKLDAYAEHKLQFWAVTAENEPSAGL 240
QY 280 LSGYFPQCLGFTPEHQDFIARDLGPTLANSTHNNVRLMLDQRLLLPHWAKVLTDP 339
DB 241 LSGYFPQCLGFTPEHQDFIARDLGPTLANSTHNNVRLMLDQRLLLPHWAKVLTDP 300
QY 340 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDRG 399
DB 301 AAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASACVSGKFWQSVRLGSDRG 360
QY 400 MOYSHSIITNLLYHVVGWTDWNLALNPEGGPNVRFVDSPIIVDTYKTFYKQPMFYHL 459
DB 361 MOYSHSIITNLLYHVVGWTDWNLALNPEGGPNVRFVDSPIIVDTYKTFYKQPMFYHL 420
QY 460 GHFSKFIPEGSQVGLVASOKNDLDAVALMHPDGSVVVYVNLNRSSKDVPLTIKDPVAGFL 519
DB 421 GHFSKFIPEGSQVGLVASOKNDLDAVALMHPDGSVVVYVNLNRSSKDVPLTIKDPVAGFL 480
QY 520 ETISPGYSIHTYLWRRQ 536
DB 481 ETISPGYSIHTYLWRRQ 497

RESULT 3

US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Fric et al.
; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULOSE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/924,440
 FILING DATE: August 27, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Christopher L. Stone
 REGISTRATION NUMBER: 35,696
 REFERENCE/DOCKET NUMBER: GC388
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 846-7555
 TELEFAX: (650) 845-6504
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 430 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-924-440-2

Query Match 3.7%; Score 105.5; DB 2; Length 430;
 Best Local Similarity 20.9%; Pred. No. 0.014;
 Matches 98; Conservative 53; Mismatches 150; Indels 169; Gaps 30;
 QY 34 VSWASGARPCPKSGY-----SSVVCVCNATYCD-----SFDPPTFPAL 73
 DB 66 VSWNSTIPPGTASFGFIASGSGEPHCTINGAPCDSESGEPGGGPGGPGDPGTPCT 125
 QY 74 GT-FSRVSTR-SGRMELSMG-PIQANHGTGILLLOPEQKQFQKVGFGGAMTDAAL 130
 DB 126 GTPVERGVQVQIGQICLDEHGNPVQLRGMS-----HGQWFDECLTD-SSL 172
 QY 131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCD-----PSIRT-----YTYAD-----T 177
 DB 173 DALAYDKADIIRLSMYIQEDGYETN-----PRGFTDRMHQILDMATARGLYIVDMHILT 238
 QY 178 PDDFQLHNSLPEEDTKLKIPLIHRALQLAORPVSLASPTPTWLKTNGAVNGKSLK 237
 DB 229 PGD-----PHYNLDRAKTFE-----AEIAQRHAS-----KTNVLY-----EIA 261
 QY 238 GQPCDIYHOTWARYFKFLDAYAEHKLQFWAFTAENEPSAGLLSG-YPPQCLGFTPEHOR 296
 DB 262 NEPGV-----SWAS-----IKSYAEVI-----PVIRQDPDSVLIIVGTRGWSLG----- 303
 QY 297 DETARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVLTDPAAKYVHGIYVHML----- 353
 DB 304 --VSESGSP-----AEIAANPVNASNIMI--AFHYAASH 334
 QY 354 --DFLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSDRGMOYSHSITNLL 411
 DB 335 RDNLY-----NALREASELFP--VFVTE--FGTETY-----TGDGANDFQADRYIDLMA 380
 QY 412 YHVVGWTDNLA-----LNP-----EGGP-----NWRNFVDS 439
 DB 381 ERKIGWTKNYSDDFRSGAVFQPGTCASGGSPWSSGSSILKASGQWVRSLQS 430

RESULT 4
 US-08-342-930-2
 Sequence 2, Application US/08342930
 Patent No. 5821084
 GENERAL INFORMATION:
 APPLICANT: OLMSTED, ELIZABETH A.
 APPLICANT: MAURO, LAURA J.
 APPLICANT: DAVIS, ALAN R.
 APPLICANT: DIXON, JACK E.
 TITLE OF INVENTION: OSTEOBLAST-TESTICULAR PROTEIN TYROSINE

TITLE OF INVENTION: PHOSPHATASE
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/342,930
 FILING DATE: 21-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: KOSKI, ANTOINETTE F.
 REGISTRATION NUMBER: 34,202
 REFERENCE/DOCKET NUMBER: 20344-20975.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1711 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-342-930-2

Query Match 3.4%; Score 97; DB 2; Length 1711;
 Best Local Similarity 19.6%; Pred. No. 1.3;
 Matches 107; Conservative 60; Mismatches 158; Indels 220; Gaps 31;
 QY 131 NILALSPPAQNLLKSYFSEEGIGYNIIRVPMASCDPSIRTYTYAD--TPDDQLHNSL 188
 DB 519 DLVDLGPDTSLSLTKSL-----VP-GSC-YTSAWAWAGNLSDSQKHSCTR 564
 QY 189 PEEDTKLKIPLIHRALQLAORPVSLASPTPTWLKTNGAVNGK----- 234
 DB 565 PAPFNL-----SLGFAQPAALKAS-WHPP-----GROAFHLRLYLRPLTL 608
 QY 235 -SLKQPGDIYHOTWARYFKFLDAYAEHKLQ---FWAVTAENEPSAGLLSGYPPQCLGF 290
 DB 609 ESEKVPREANQFSNAQ-----LTAGECFQVQLSTLWG--SERSSA-----NATGW 653
 QY 291 TPEQRDFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKV-----VLTDEPAK 342
 DB 654 TPTS-----APTLVNVIS-----DAPQLQVSWAHVPGGSRGYQVTLQESTR 696
 QY 343 YVHGI-----AVHWYLDLAPAKATLGETHRLFPNTMLFASE- 379
 DB 697 TATSMGPKEDGTSFLGTPGTYKVEVISWAGPLYTAAANVSAWYPLPLNELLVSMQA 756
 QY 380 -----AC-----VGSKFWEQSVRLGS---WDRGMOYSHSI----- 406
 DB 757 GSAVVNLWPSGLPGQACHAQLSDGHLSEWPKLQGFMLRDLTPTGHTISMSVRCR 816
 QY 407 -----ITNLLYH-----VVGWT-----DNLA-----LNDEGPN 431
 DB 817 AGPLOASTHLVLSVEPGVPEVLCPEATYLAUNWTMPAGDVDCVLVVERLVPGGTH 876
 QY 432 W---VRNFVDSPIIVDITKDTYKOPMYHGHFSKIPGEGSORGLVASQKNDLDAVAL 488
 DB 877 FVQVNTSGDALLPLNLMPTTSYLSLTV-LGRNSRW-----SRVSLVCS-----TSEA 926
 QY 489 MHP-----DCSAVVVNLNRSSKQVPLTIKDPV---GFLETIS-----PGYSI 528
 DB 927 WHPELAEPQVQLGTGMGVTVNRGMFGKD-----DQIQWYGIATINMTLAQPSREA 980

QY 529 HTYLW 533
Db 981 INYTW 985

RESULT 5

US-08-924-759-24
; Sequence 24, Application US/08924759

; Patent No. 5962229

; GENERAL INFORMATION:

; APPLICANT: MCGONIGLE, BRIAN

; APPLICANT: O'KEEFE, DANIEL

; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 19898

; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.50 INCH

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95

; SOFTWARE: MICROSOFT WORD VERSION 7.0A

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/924,759

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: FLOYD, LINDA AXAMETHY

; REGISTRATION NUMBER: 33,692

; REFERENCE/DOCKET NUMBER: CL-1128

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 302-892-8112

; TELEFAX: 302-773-0164

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 180 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; TISSUE TYPE: MAIZE

; IMMEDIATE SOURCE:

; LIBRARY: CS1.PK0059.E2

; US-08-924-759-24

Query Match 3.3%; Score 94.5; DB 2; Length 180;

Best Local Similarity 25.4%; Pred. No. 0.047; 55; Indels 29; Gaps 7;

Matches 36; Conservative 22; Mismatches 22; Indels 29; Gaps 7;

QY 143 LKSYFSEEGIGYNIIRVPMASCDIFSIRTYADTPDDFQLHNFSLPERDTKLKIPLIHR 202

Db 9 LLGSWYSP-----YVIRAKVA---LGLKGLSYEFVEEDLS-----RKSDLLKLNPNVHR 54

QY 203 ALQL----AORPV---SLLASPTWSTPTWLKTNAGVNGKSLKGGPDIYHOTWARYFVKFLD 257

Db 55 KVPVLVHGGPRVCESLVIQYVDETWAGTGTP-----LPADAYDRAMARFWAAAYVD 106

QY 258 --AYAETHKLOFWAVTAENEPSA 277

Db 107 DKFYKEWNRFLFWSTTAEKAAEA 128

RESULT 6

US-09-248-335-24

; Sequence 24, Application US/09248335

; Patent No. 6096504

; GENERAL INFORMATION:

; APPLICANT: MCGONIGLE, BRIAN

; APPLICANT: O'KEEFE, DANIEL

; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES

; FILE REFERENCE: CL-1128-A

; CURRENT APPLICATION NUMBER: US/09/248,335

; CURRENT FILING DATE: 1999-02-10

; EARLIER APPLICATION NUMBER: 08/924,759

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: Microsoft Word Version 7.0A

; SEQ ID NO 24

; LENGTH: 180

; TYPE: PRT

; ORGANISM: maize

; US-09-248-335-24

Query Match 3.3%; Score 94.5; DB 3; Length 180;

Best Local Similarity 25.4%; Pred. No. 0.047; 55; Indels 29; Gaps 7;

Matches 36; Conservative 22; Mismatches 22; Indels 29; Gaps 7;

QY 143 LKSYFSEEGIGYNIIRVPMASCDIFSIRTYADTPDDFQLHNFSLPERDTKLKIPLIHR 202

Db 9 LLGSWYSP-----YVIRAKVA---LGLKGLSYEFVEEDLS-----RKSDLLKLNPNVHR 54

QY 203 ALQL----AORPV---SLLASPTWSTPTWLKTNAGVNGKSLKGGPDIYHOTWARYFVKFLD 257

Db 55 KVPVLVHGGPRVCESLVIQYVDETWAGTGTP-----LPADAYDRAMARFWAAAYVD 106

QY 258 --AYAETHKLOFWAVTAENEPSA 277

Db 107 DKFYKEWNRFLFWSTTAEKAAEA 128

RESULT 7

US-08-471-119A-2

; Sequence 2, Application US/084711119A

; Patent No. 5827706

; GENERAL INFORMATION:

; APPLICANT: Leitner, Ernst

; APPLICANT: Schneider, Elisabeth

; APPLICANT: Schoergendorfer, Kurt

; APPLICANT: Weber, Gerhard

; TITLE OF INVENTION: Cyclosporin Synthetase

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5827706artis Corporation

; STREET: 59 Route 10

; CITY: East Hanover

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07936

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/471,119A

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Kassenoff, Melvyn

; REGISTRATION NUMBER: 26,389

; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 503 8474

; TELEFAX: 201 503 8807

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15281 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match 3.2%; Score 93; DB 2; Length 15281;
Best Local Similarity 19.7%; Pred. No. 1.7e+02;
Matches 122; Conservative 70; Mismatches 235; Indels 192; Gaps 28;

QY 9 EEC-----PKPLSRVSIM--AGSLTGILLQAVSWASGARP---CIPKSGYSSVVCVN 58
DB 7502 QECRLSVLDQPLTPISVLPFSNAISNLESLLLEMPSTSDYPRDRTVVDLFRQAACPD 7561
QY 59 ATYCDSDPPTPALGTSTSRVSTSRGRMELSMGPQOANHGTGTLGLLTLQPEQKFKVK 118
DB 7562 IAVKSSSLTYAQLDE---QSDRAVAWLHERHMPAES-----LVGLSPRSCETIIA 7611
QY 119 GFGAMTDAANILALSPPA-----ONLLKSYFSEEGICYNIIRVMASCDF 167
DB 7612 YFGIMKANLAYLPDYADARLAALDITVEGERLLLG---AGVPQCIQIPRLSTAY 7667
QY 168 SIRTYYADPDQFQHNLSPEED-TKLK-----IPLIIRALQLAQR--- 209
DB 7668 IAEALSHATTVD-----VTSIPQSATSIAKVIPTSGTCKPKGMIEHGVRLVRDTN 7722
QY 210 -----PVSLIAS-PWTSPTWLKINGAVNGKSLGKQPGDIYHOTWARYFVKFLD 257
DB 7723 VNVFPGSGALPVSHFNSLNDAAATYITAVLNG---GTVCVDRDT-----MLD 7770
QY 258 AYAEHLQFWATVNAEPNEPSAGLLSGYPPQCLGTFPEHQRFIARDLGPPLANSHTHNVRL 317
DB 7771 IAAIINS-----TFKENVRAAFTPAFLKQCLAETPE---LVA-----NLEI 7809
QY 318 LMLDDORILLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKA-----TLGETHRLFP 371
DB 7810 LHAGDRL-----DPGDAN-----LAGTKAGGIENVLGHTE--- 7841
QY 372 NTMLFASACVSKFEQSGVSLGSDRCMOYSHSITNLLYHVVGWTDNLANLNEP--- 428
DB 7842 NTAYSFYFVPGEEFVNGVPG---RCISNSHAYI-----IDRQKLVPAVWG 7888
QY 429 -----GPNVRFVDSPIIVDITKDFY-----KQPMFVHLGHFSKFTPEGSQ----- 471
DB 7889 ELITGDGVARGYTDSA-----LNKDRFVYIDINGKSTWSTYTGDKARYRPRDQLEFFGR 7944
QY 472 -----RVGLVASOKNDLDAVALMHPDGSVAVVVLNRSSKSDVPLTIKDPVAG----- 517
DB 7945 MDQMVKIRGVRIEPEGEVELTLDHKSVLAATVVVVRPPNGDPEMIAFITIDAEDDVQTHK 8004
QY 518 -----FLETISPGYSIHTYL 532
DB 8005 AIYKHLQGLPAYMTPSHL 8023

RESULT 8
US-08-971-188-5
; Sequence 5, Application US/08971188
; Patent No. 6326165
; GENERAL INFORMATION:
; APPLICANT: Wilson, Thomas G.
; APPLICANT: Helrich, Julia N.
; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO
; TITLE OF INVENTION: SCREEN POTENTIAL INSECTICIDES
; FILE REFERENCE: 082584/102
; CURRENT APPLICATION NUMBER: US/08/971,188
; CURRENT FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 08/843,205
; EARLIER FILING DATE: 1997-04-14
; NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 5
LENGTH: 716
TYPE: PRT
ORGANISM: Met-JHR
FEATURE:
OTHER INFORMATION: xaa at position 103 may be Gly or Arg
US-08-971-188-5

Query Match 3.2%; Score 92; DB 4; Length 716;
Best Local Similarity 27.7%; Pred. No. 1;
Matches 31; Conservative 15; Mismatches 34; Indels 32; Gaps 4;

QY 449 TFYQPMFYHLGHFSKFIPEGS-----QRVGLVASOKND---LDAVALMHPDGSVAVV 499
DB 392 TYIROPEPYOLEYHTRHLIDGSIIDCDQRIGLVAGYMKDEVRLNSPFCFHLDDVRWVIV 451
QY 500 -----LNRSSKSDVPLTIKDPVAGFLETISPGYSIHTYL 532
DB 452 ALROMYDCNSDYGSCYRLLSRNGRFIYLTK-----GFLEVDGRGSKNVHSFL 499

RESULT 9
US-08-971-188-11
; Sequence 11, Application US/08971188
; Patent No. 6326165
; GENERAL INFORMATION:
; APPLICANT: Wilson, Thomas G.
; APPLICANT: Helrich, Julia N.
; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO
; TITLE OF INVENTION: SCREEN POTENTIAL INSECTICIDES
; FILE REFERENCE: 082584/102
; CURRENT APPLICATION NUMBER: US/08/971,188
; CURRENT FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 08/843,205
; EARLIER FILING DATE: 1997-04-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 11
LENGTH: 716
TYPE: PRT
ORGANISM: Met-JHR
US-08-971-188-11

Query Match 3.2%; Score 92; DB 4; Length 716;
Best Local Similarity 27.7%; Pred. No. 1;
Matches 31; Conservative 15; Mismatches 34; Indels 32; Gaps 4;

QY 449 TFYQPMFYHLGHFSKFIPEGS-----QRVGLVASOKND---LDAVALMHPDGSVAVV 499
DB 392 TYIROPEPYOLEYHTRHLIDGSIIDCDQRIGLVAGYMKDEVRLNSPFCFHLDDVRWVIV 451
QY 500 -----LNRSSKSDVPLTIKDPVAGFLETISPGYSIHTYL 532
DB 452 ALROMYDCNSDYGSCYRLLSRNGRFIYLTK-----GFLEVDGRGSKNVHSFL 499

RESULT 10
US-09-801-052-4
; Sequence 4, Application US/09801052
; Patent No. 6368842
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen
; APPLICANT: Helrich, Julia N.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO01045
; CURRENT APPLICATION NUMBER: US/09/801,052
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 437

Sequence 8, Application US/09052469
 Patent No. 6380360
 GENERAL INFORMATION:
 APPLICANT: Harris et al.
 TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: One Financial Center
 CITY: Boston
 STATE: MA
 COUNTRY: US
 ZIP: 02111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk, 3.50 inch
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WordPerfect 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/052,469
 FILING DATE: Concurrently herewith
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/422,582
 FILING DATE: 14-April-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9507766.5
 FILING DATE: 13-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9411900.5
 FILING DATE: 14-JUN-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB PCT/GB94/02822
 FILING DATE: 23-DEC-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9326470.3
 FILING DATE: 24-DEC-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams, Ph.D., Kathleen M.
 REGISTRATION NUMBER: 34,360
 REFERENCE/DOCKET NUMBER: 3265/74165
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 345-9100
 TELEFAX: (617) 345-9111
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4302 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-052-469-8

Query Match 3.1%; Score 88.5; DB 4; Length 4302;
 Best Local Similarity 22.0%; Pred. No. 57;
 Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;
 Qy 269 VTAEPEPSAG-----LLSGYPPQCLG---FTPEHQRDFA----- 300
 Db 1611 VTAEVNGSAQDSIFVYVLIQLEG--LQVGGGRYFPTNHTVQLQAVVRDGTNVSYSWTA 1668
 Qy 301 -RDLGPTLANS-----THNNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGI 347
 Db 1669 WRDRGPAAGSGKGFSLTVLEAGTYH---VOLRATNMLGSANADCTMDFE----- 1716
 Qy 348 AVHWYLDLAPAKATIGETHRLFPNTMLFASEACVSK--FWQSVRLG-SWDRGMQY-S 403
 Db 1717 PVGWLMTASPNPAV-----NTSVTLSAELAGSGVYVYTWLSLEGLSWETSEPTT 1768
 Qy 404 HSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDIKDFYKQPM----- 455
 Db 1769 HSFPTGCL-HLVMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1810
 Qy 456 -----FYLGHFSKF-----IPEGSQRVGLVSAQKNDLDAVALMHPDGA 495

Db 1811 EPGGSEFVAAGSSVFPWGLATGNTVSWCHAVPGSSKRG-----PHYTMVFPDAGT 1861
 Qy 496 VVVVLRNS-----SKDVPITIKDPVAGFL-----ETISPGYSIH 529
 Db 1862 FSIRLNASNAWSWATYNTLTAEEPIVGLVIMASSKVWAPGQLVH 1906
 RESULT 14
 US-08-460-751-2
 ; Sequence 2, Application US/08460751
 ; Patent No. 5891628
 ; GENERAL INFORMATION:
 ; APPLICANT: Readers, Stephen
 ; APPLICANT: Schneider, Michael
 ; APPLICANT: Glucksmann, Sandra
 ; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
 ; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/460,751
 ; FILING DATE: 02-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/413,580
 ; FILING DATE: 03-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7638-005
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4303 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-460-751-2

Query Match 3.1%; Score 88.5; DB 2; Length 4303;
 Best Local Similarity 22.0%; Pred. No. 57;
 Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;
 Qy 269 VTAEPEPSAG-----LLSGYPPQCLG---FTPEHQRDFA----- 300
 Db 1611 VTAEVNGSAQDSIFVYVLIQLEG--LQVGGGRYFPTNHTVQLQAVVRDGTNVSYSWTA 1668
 Qy 301 -RDLGPTLANS-----THNNVRLMLDDQRLLLPHWAKVVLTDPEAAKYVHGI 347
 Db 1669 WRDRGPAAGSGKGFSLTVLEAGTYH---VOLRATNMLGSANADCTMDFE----- 1716
 Qy 348 AVHWYLDLAPAKATIGETHRLFPNTMLFASEACVSK--FWQSVRLG-SWDRGMQY-S 403
 Db 1717 PVGWLMTASPNPAV-----NTSVTLSAELAGSGVYVYTWLSLEGLSWETSEPTT 1768
 Qy 404 HSIITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDIKDFYKQPM----- 455
 Db 1769 HSFPTGCL-HLVMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1810

QY 456 -----FYLGHFSKF-----IPEGQRVGLVASOKNDLDAVALMHPDGS 495
Db 1811 EPGGSFVAAGSSVPFWGLATGNTVSWCWA VPGGSSKRG-----PHTVMVFPDAGT 1861
QY 496 VVVVLNRS-----SKDVPITIKDPVAGFL-----ETISPGYSIH 529
Db 1862 FSIRLNASNAVSWVSATYNLTAEPIVGLVLWASSKVYVAPGQLVH 1906

RESULT 15
US-09-052-469-6
; Sequence 6, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-469-6

Query Match 3.1%; Score 88.5; DB 4; Length 4339;
Best Local Similarity 22.0%; Pred. No. 58;
Matches 76; Conservative 34; Mismatches 102; Indels 133; Gaps 20;
QY 269 VTAENEPSAG-----LLSGYFPFOCLG-----FTPEHQDFIA----- 300
Db 1568 VTAENEVSADSIFFVYVLIQIEG--LQVVGGRYFPTNHTVOLQAVVRDGTNVSYSWTA 1625
QY 301 -RDIGPTLANS-----THHNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGI 347

Db 1626 WRDRGPALAGSGKGFSLTVLEAGTYH-----VOLRATNMLGSAWADCTMDFVE----- 1673
QY 348 AVHWYLDLFLAPAKATLGETHRLFPNTMLFASEACVGSK--FWEQSVRLG--SWDRGMQY-S 403
Db 1674 PVGWLMTASPNPAAY-----NTSVTLSAELAGSGGVYVYVWLSLEGLSWETSEPTT 1725
QY 404 HSIITNLLYHVYGTWDMNLALNPEGGPNNVRNFVDSPIIVDITKDTFYKOPM----- 455
Db 1726 HSFPTPGL-HLYTMT-----AGNPLGSAN-----ATVEVDV-----QVPVSGLSIRAS 1767
QY 456 -----FYHLGHFSKF-----IPEGQRVGLVASOKNDLDAVALMHPDGS 495
Db 1768 EPGGSFVAAGSSVPFWGLATGNTVSWCWA VPGGSSKRG-----PHTVMVFPDAGT 1818
QY 496 VVVVLNRS-----SKDVPITIKDPVAGFL-----ETISPGYSIH 529
Db 1819 FSIRLNASNAVSWVSATYNLTAEPIVGLVLWASSKVYVAPGQLVH 1863

Search completed: February 20, 2003, 11:21:03
Job time : 47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:07:51 ; Search time 39 Seconds

(without alignments)
1831.342 Million cell updates/sec

Title: US-10-024-197-25

Perfect score: 2865

Sequence: 1 MFSPSPRECPKPLRSVSI.....GFLETISPGSYHTLWRRQ 536

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2861	99.9	546	AAW18237	Human glucocerebro
2	2854	99.6	536	AAW18237	Protein encoded by
3	2854	99.6	536	AAW18237	Human glucocerebro
4	2832	98.8	536	AAW18237	Human glucocerebro
5	2762	96.4	516	AAW18237	Human lysosomal gl
6	2762	96.4	516	AAW18237	Human lysosomal gl
7	2762	96.4	516	AAW18237	Amino acid sequenc
8	2757	96.2	516	AAW18237	Human glucocerebro
9	2757	96.2	516	AAW18237	Amino acid sequenc
10	2736	95.5	516	AAW18237	Recombinant human

11	2675	93.4	592	22	AAU05698	Human glucocerebro
12	2665	93.0	498	17	AAW07885	Remodelled, recomb
13	2664	93.0	497	22	AAU05696	Human Mature gluco
14	2664	93.0	497	23	AAU06020	Human glucocerebro
15	2659	92.8	497	22	AAU05688	Human glucocerebro
16	2659	92.8	497	22	AAU05680	Human glucocerebro
17	2659	92.8	497	22	AAU05684	Human glucocerebro
18	2658	92.8	497	22	AAU05699	Human glucocerebro
19	2656	92.7	497	22	AAU05667	Human glucocerebro
20	2656	92.7	497	22	AAU05676	Human glucocerebro
21	2655	92.7	497	22	AAU05678	Human glucocerebro
22	2655	92.7	497	22	AAU05681	Human glucocerebro
23	2655	92.7	497	22	AAU05685	Human glucocerebro
24	2655	92.7	497	22	AAU05692	Human glucocerebro
25	2655	92.7	497	22	AAU05654	Human glucocerebro
26	2654	92.6	497	22	AAU05654	Human glucocerebro
27	2654	92.6	497	22	AAU05656	Human glucocerebro
28	2654	92.6	497	22	AAU05662	Human glucocerebro
29	2654	92.6	497	22	AAU05670	Human glucocerebro
30	2654	92.6	497	22	AAU05671	Human glucocerebro
31	2654	92.6	497	22	AAU05675	Human glucocerebro
32	2654	92.6	497	22	AAU05688	Human glucocerebro
33	2654	92.6	497	22	AAU05653	Human glucocerebro
34	2653	92.6	497	22	AAU05655	Human glucocerebro
35	2653	92.6	497	22	AAU05661	Human glucocerebro
36	2653	92.6	497	22	AAU05665	Human glucocerebro
37	2653	92.6	497	22	AAU05669	Human glucocerebro
38	2653	92.6	497	22	AAU05672	Human glucocerebro
39	2653	92.6	497	22	AAU05673	Human glucocerebro
40	2653	92.6	497	22	AAU05674	Human glucocerebro
41	2653	92.6	497	22	AAU05677	Human glucocerebro
42	2653	92.6	497	22	AAU05679	Human glucocerebro
43	2653	92.6	497	22	AAU05689	Human glucocerebro
44	2653	92.6	497	22	AAU05691	Human glucocerebro
45	2653	92.6	497	22	AAU05691	Human glucocerebro

ALIGNMENTS

RESULT 1
AAW18237
ID AAW18237 standard; Protein; 546 AA.
XX AAW18237;
AC AAW18237;
XX 01-OCT-1997 (first entry)
XX Human glucocerebrosidase-FLAG fusion.
XX Lysosomal enzyme; lysosome; transgenic plant; glucocerebrosidase;
XX enzyme replacement therapy; Gaucher disease.
XX Chimaeric Homo sapiens;
OS Chimaeric synthetic.
XX Key Location/Qualifiers
FH Protein 1..536
FT Peptide /label= hCG
FT Peptide 539..546
FT Peptide /label= FLAG
FT Misc-difference 89
FT /note= "encoded by GGG"

PA (VIRG) VIRGINIA TECH INTELLECTUAL PTY INC.
 XX Cramer CL, Olshi KK, Radin DN, Weissenborn DL;
 XX WPI; 1997-202248/18.
 DR N-PSDB; AAT71753.
 XX
 XX Production of enzymatically active (modified) lysosomal enzyme in
 PT transgenic plants - useful in treatment of lysosomal storage
 PT disorders
 XX
 XX Example 1; Page 62-64; 11pp; English.
 XX
 XX A novel fusion protein (AAW18237) comprises human glucocerebrosidase
 CC (hgc) and a C-terminal FLAG epitope (see also AAW18236). A DNA
 CC construct (see also AAT71753) encoding the fusion is placed under
 CC control of the inducible MeGA promoter (see also AAT71752) in a
 CC plant expression system to provide production of recombinant hgc
 CC in transgenic (esp. tobacco) plants. The plant expression system
 CC provides for post-translational modification and processing to
 CC produce enzymatically active hgc, which can be harvested for use in
 CC Gaucher disease enzyme replacement therapy.
 XX
 XX Sequence 546 AA;
 SQ
 Query Match 99.9%; Score 2861; DB 18; Length 546;
 Best Local Similarity 99.6%; Pred. No. 2.1e-290;
 Matches 534; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEFSSPRECECPKLSRVSMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 DB 1 MEFSSPRECECPKLSRVSMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 QY 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTLPQEQKFKVKG 120
 DB 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTLPQEQKFKVKG 120
 QY 121 GGAMTDAALNIALSPQAQNLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
 DB 121 GGAMTDAALNIALSPQAQNLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
 QY 181 FQHNFSLEPEDTKLPIHLRALQAPVSLASPTWTLKTNAGVNGKSLGQ 240
 DB 181 FQHNFSLEPEDTKLPIHLRALQAPVSLASPTWTLKTNAGVNGKSLGQ 240
 QY 241 GDYHQTWARYFVKFLDAYAEHLQFVAETAENEPSAGLLSGYFPQCIGFTPEHQ 300
 DB 241 GDYHQTWARYFVKFLDAYAEHLQFVAETAENEPSAGLLSGYFPQCIGFTPEHQ 300
 QY 301 RDLGPTLANSTHNNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGIAVHWYLD 360
 DB 301 RDLGPTLANSTHNNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGIAVHWYLD 360
 QY 361 ATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDWDRGMQYSHSITNLLYH 420
 DB 361 ATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDWDRGMQYSHSITNLLYH 420
 QY 421 NLALNPEGGPNVNFVDSPIIVDTKDTFYKQPMFYHLGHSKFIPEGSQ 480
 DB 421 NLALNPEGGPNVNFVDSPIIVDTKDTFYKQPMFYHLGHSKFIPEGSQ 480
 QY 481 NDLDAVALMHPDGSVAVVNLNRSSKDVPLTIKDPVAGFLETISP 536
 DB 481 NDLDAVALMHPDGSVAVVNLNRSSKDVPLTIKDPVAGFLETISP 536
 RESULT 2
 AAB61495
 ID AAB61495 standard; Protein: 536 AA.
 XX
 AC AAB61495;
 XX
 DT 04-APR-2001 (first entry)

XX Protein encoded by unmodified Gcc DNA.
 DE
 XX Glucocerebrosidase; Gcc; splice; gene therapy; enzyme therapy;
 KW Gaucher.
 XX
 XX Unidentified.
 OS
 XX CA2272055-A1.
 PN
 XX 02-DEC-2000.
 PD
 XX 02-JUN-1999; 99CA-2272055.
 PF
 XX 02-JUN-1999; 99CA-2272055.
 PR
 XX (HSCR-) HSC RES & DEV LP.
 PA
 XX Mahuran DJ, Callahan JW, Clarke JTR;
 PI
 XX WPI; 2001-160192/17.
 DR
 XX New glucocerebrosidase (Gcc) DNA molecule modified in at least one
 PT nucleotide, useful in gene therapy or enzyme replacement therapy,
 PT particularly in treating Gaucher disease
 PT
 PS Disclosure; Fig 4; 37pp; English.
 PS
 XX The present invention relates to glucocerebrosidase (Gcc) DNA
 CC with a modification in at least one nucleotide that disrupts a
 CC splicing consensus sequence and prevents splicing of mRNA produced
 CC from the DNA molecule, while preserving the ability of the DNA to
 CC express active Gcc. The glucocerebrosidase (Gcc) DNA molecule and
 CC the vector comprising the DNA are useful in gene therapy or enzyme
 CC replacement therapy, particularly in treating Gaucher disease. The
 CC cells transfected or transduced by the vectors are useful as research
 CC tools.
 CC
 SQ Sequence 536 AA;
 Query Match 99.6%; Score 2854; DB 22; Length 536;
 Best Local Similarity 99.6%; Pred. No. 1.1e-289;
 Matches 534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NEFSSPRECECPKLSRVSMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 DB 1 NEFSSPRECECPKLSRVSMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 QY 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTLPQEQKFKVKG 120
 DB 61 YCDSFDPTFPALGTFSTRSGRRMELSMGPIQANHTGTGLLLTLPQEQKFKVKG 120
 QY 121 GGAMTDAALNIALSPQAQNLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
 DB 121 GGAMTDAALNIALSPQAQNLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDD 180
 QY 181 FQHNFSLEPEDTKLPIHLRALQAPVSLASPTWTLKTNAGVNGKSLGQ 240
 DB 181 FQHNFSLEPEDTKLPIHLRALQAPVSLASPTWTLKTNAGVNGKSLGQ 240
 QY 241 GDYHQTWARYFVKFLDAYAEHLQFVAETAENEPSAGLLSGYFPQCIGFTPEHQ 300
 DB 241 GDYHQTWARYFVKFLDAYAEHLQFVAETAENEPSAGLLSGYFPQCIGFTPEHQ 300
 QY 301 RDLGPTLANSTHNNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGIAVHWYLD 360
 DB 301 RDLGPTLANSTHNNVRLMLDDQRLPHWAKVVLTDPEAAKYVHGIAVHWYLD 360
 QY 361 ATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDWDRGMQYSHSITNLLYH 420
 DB 361 ATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDWDRGMQYSHSITNLLYH 420
 QY 421 NLALNPEGGPNVNFVDSPIIVDTKDTFYKQPMFYHLGHSKFIPEGSQ 480
 DB 421 NLALNPEGGPNVNFVDSPIIVDTKDTFYKQPMFYHLGHSKFIPEGSQ 480

DB 421 NLALNPEGGPNVRFVDSPIIIVDTIKDFYKQPMFYHLGHFSKFIPEGSORVGLVASOK 480
 QY 481 NDLDVALMHDPDGSVAVVVLNRSKDVPLTIKDPAGVPLETISPCYSIHTYLMRRQ 536
 DB 481 NDLDVALMHDPDGSVAVVVLNRSKDVPLTIKDPAGVPLETISPCYSIHTYLMRRQ 536

RESULT 3
 ID AAB61496 standard; Protein; 536 AA.
 AC AAB61496;
 DT 04-APR-2001 (first entry)
 DE protein encoded by modified Gcc DNA.
 XX Glucocerebrosidase; Gcc; splice; gene therapy; enzyme therapy;
 KW Gaucher.
 XX Unidentified.
 XX CA2272055-A1.
 PN 02-DEC-2000.
 PD 02-JUN-1999; 99CA-2272055.
 PF 02-JUN-1999; 99CA-2272055.
 PR 02-JUN-1999; 99CA-2272055.
 XX (HSCR-) HSC RES & DEV LP.
 XX Mahuran DJ, Callahan JW, Clarke JTR;
 PI WPI; 2001-160192/17.
 DR New glucocerebrosidase (gcc) DNA molecule modified in at least one
 XX nucleotide, useful in gene therapy or enzyme replacement therapy,
 PT particularly in treating Gaucher disease -
 PS Disclosure; Fig 4; 37pp; English.
 XX The present invention relates to glucocerebrosidase (gcc) DNA
 CC with a modification in at least one nucleotide that disrupts a
 CC splicing consensus sequence and prevents splicing of mRNA produced
 CC from the DNA molecule, while preserving the ability of the DNA to
 CC express active gcc. The glucocerebrosidase (gcc) DNA molecule and
 CC the vector comprising the DNA are useful in gene therapy or enzyme
 CC replacement therapy, particularly in treating Gaucher disease. The
 CC cells transfected or transduced by the vectors are useful as research
 CC tools.

Sequence 536 AA;
 Query Match 99.6%; Score 2854; DB 22; Length 536;
 Best Local Similarity 99.6%; Pred. No. 1.1e-289;
 Matches 534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MEFSFSPRECPKPLSRYSIMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 DB 1 MEFSFSPRECPKPLSRYSIMAGSLTGLLLQAVSWASGARPCIPKSGYSSVVCVNCAT 60
 QY 61 YCDSFDPPTFPALGTFSRYSTRGRMELSMGPQIANHTGTGLLTLPQKOKVKCF 120
 DB 61 YCDSFDPPTFPALGTFSRYSTRGRMELSMGPQIANHTGTGLLTLPQKOKVKCF 120
 QY 121 GGAWTDAALNIALSPQAQNLKLSFSEEGYNYIIRVPMASCDFTIRTYADTPDD 180
 DB 121 GGAWTDAALNIALSPQAQNLKLSFSEEGYNYIIRVPMASCDFTIRTYADTPDD 180
 QY 181 FOLHNSFLPEEDTKLKIPLIHRALQALQRPVSLASPTWTKTNGAVNGKSLKGQP 240

DB 181 FOLHNSFLPEEDTKLKIPLIHRALQALQRPVSLASPTWTKTNGAVNGKSLKGQP 240
 QY 241 GDIIYHQTWARYFVFLDAYAEHKLQFWAVTAENPSAGLLSGYPFQCLGTFPEHORDFTA 300
 DB 241 GDIIYHQTWARYFVFLDAYAEHKLQFWAVTAENPSAGLLSGYPFQCLGTFPEHORDFTA 300
 QY 301 RDLGPTLANSTHNNRLLMLDDQRLLLPHWAKVLTDPFAAKYVHGIHVLDFLAPAK 360
 DB 301 RDLGPTLANSTHNNRLLMLDDQRLLLPHWAKVLTDPFAAKYVHGIHVLDFLAPAK 360
 QY 361 ATLGETHRLFPNTMLFASEACVGSKEWQSVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
 DB 361 ATLGETHRLFPNTMLFASEACVGSKEWQSVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
 QY 421 NLALNPEGGPNVRFVDSPIIIVDTIKDFYKQPMFYHLGHFSKFIPEGSORVGLVASOK 480
 DB 421 NLALNPEGGPNVRFVDSPIIIVDTIKDFYKQPMFYHLGHFSKFIPEGSORVGLVASOK 480
 QY 481 NDLDVALMHDPDGSVAVVVLNRSKDVPLTIKDPAGVPLETISPCYSIHTYLMRRQ 536
 DB 481 NDLDVALMHDPDGSVAVVVLNRSKDVPLTIKDPAGVPLETISPCYSIHTYLMRRQ 536

RESULT 4
 ID AAE02446 standard; Protein; 536 AA.
 AC AAE02446;
 DT 10-AUG-2001 (first entry)
 DE Human glucocerebrosidase (GC) #2.
 XX Human; adeno-associated viral expression vector; AAV; gene therapy;
 KW lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
 KW Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG;
 XX glucocerebrosidase; GC; Gaucher's disease.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH 1..39 Peptide
 FT /label= Signal_peptide
 FT Protein 40..536
 FT /note= "Human mature glucocerebrosidase (GC)"
 XX WO200136603-A2.
 XX 25-MAY-2001.
 XX 17-NOV-2000; 2000WO-US31688.
 XX 17-NOV-1999; 99US-0166097.
 XX 30-JUN-2000; 2000US-0215430.
 XX (AVIG-) AVIGEN INC.
 XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
 XX Podsakoff G, Watson G, Couto LB, Yang B;
 XX WPI; 2001-343814/36.
 XX N-PSDB; AAD06389.
 XX Use of recombinant adeno-associated virus, comprising gene encoding a
 XX protein defective or missing in lysosomal storage disease, in the
 XX manufacture of a medicament for treating the lysosomal storage disease
 XX -
 XX Disclosure; Page 94-96; 97pp; English.
 XX The present invention relates to recombinant adeno-associated virus (AAV)
 XX expression vectors and virions, which include genes coding for enzymes
 XX defective or missing in lysosomal storage disease (LSD). AAV is useful

CC in the manufacture of a medicament for treating lysosomal storage
 CC disease e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome)
 CC is due to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS)
 CC which aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in
 CC gene therapy. The present sequence is human glucocerebrosidase (GC).
 CC Mutations in GC gene leads to Gaucher's disease, a lysosomal
 CC storage disease. This sequence is used in AAV constructs.

XX
 SQ Sequence 536 AA;

Query Match 98.8%; Score 2832; DB 22; Length 536;
 Best Local Similarity 98.9%; Pred. No. 2.3e-287;
 Matches 530; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MEFSRRECPKPLSRVSIMAGSLTGLLLQAVSWASGARPCPKSFSGYSSVVCVNCAT 60
 DB 1 MEFSRRECPKPLNDPEGMAGSLTGLLLQAVSWASGARPCPKSFSGYSSVVCVNCAT 60
 QY 61 YCDSFDPPTFPALGTFSTRSGRRMELSMGPIQANHGTGTLTLOPEQKFKVKG 120
 DB 61 YCDSFDPPTFPALGTFSTRSGRRMELSMGPIQANHGTGTLTLOPEQKFKVKG 120
 QY 121 GGAMTAAALNIALSPPAQNLLKSYSEEGIGYNIIRVPMASCDFSIRTYTYADTDD 180
 DB 121 GGAMTAAALNIALSPPAQNLLKSYSEEGIGYNIIRVPMASCDFSIRTYTYADTDD 180
 QY 181 FOLHNFSLPEEDTKLPIHRAALQALQAPVSLASPTPTWLTNGAVNGKSLGQ 240
 DB 181 FOLHNFSLPEEDTKLPIHRAALQALQAPVSLASPTPTWLTNGAVNGKSLGQ 240
 QY 241 GDYIHWYVYVFKFLDAYAEHKLQFVAWTAENEPSAGLLSGYPCQCLGTFPEHORD 300
 DB 241 GDYIHWYVYVFKFLDAYAEHKLQFVAWTAENEPSAGLLSGYPCQCLGTFPEHORD 300
 QY 301 RDLGPTLANSTHNNVLLMDQRLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
 DB 301 RDLGPTLANSTHNNVLLMDQRLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
 QY 361 ATLGETHRLFPNTMLFASACVGSQVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
 DB 361 ATLGETHRLFPNTMLFASACVGSQVRLGSDRGMOYSHSIITNLLYHVVGWTDW 420
 QY 421 NLALNPEGGPNVNRVDSPIIVDTKDTFYKQPMFYHLGHFSKFIPEGSRVGLVASQ 480
 DB 421 NLALNPEGGPNVNRVDSPIIVDTKDTFYKQPMFYHLGHFSKFIPEGSRVGLVASQ 480
 QY 481 NDLDAVALMHPDGSAAVVVNLNRSSKDVPLTIKDPVAGFLETISPGYSIHTYLWRRQ 536
 DB 481 NDLDAVALMHPDGSAAVVVNLNRSSKDVPLTIKDPVAGFLETISPGYSIHTYLWRRQ 536

RESULT 5
 AAY01595
 ID AAY01595 standard; Protein; 516 AA.
 XX
 AC AAY01595;
 XX
 XX
 17-JUN-1999 (first entry)
 XX
 DE Human lysosomal glucocerebrosidase protein.
 XX
 KW Human lysosomal glucocerebrosidase; glycosylated; Gaucher's disease.
 XX
 OS Homo sapiens.
 XX
 PN US5879680-A.
 XX
 PD 09-MAR-1999.
 XX
 PF 26-MAY-1995; 95US-0452398.
 XX
 PR 23-DEC-1987; 87US-0137796.
 PR 05-FEB-1990; 90US-0474307.

PR 06-AUG-1992; 92US-0925333.
 PR 13-JAN-1994; 94US-0186256.
 PR 26-MAY-1995; 95US-0452398.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Eliason WK, Ginns EJ, LaMarca ME, Martin B, Maysak KA;
 PI WPI; 1999-203905/17.
 DR N-PSDB; AAX26682.
 XX
 XX New compositions containing glucocerebrosidase - comprising
 PT glycosylated recombinantly-produced human glucocerebrosidase, used
 PT for treating Gaucher's disease
 XX
 XX Claim 1; Fig 1A-D; 10pp; English.
 PS
 XX The present sequence represents a human lysosomal glucocerebrosidase.
 CC The specification describes a composition containing glycosylated
 CC recombinantly produced human glucocerebrosidase which can be used
 CC for treating Gaucher's disease.
 XX
 SQ Sequence 516 AA;

Query Match 96.4%; Score 2762; DB 20; Length 516;
 Best Local Similarity 100.0%; Pred. No. 4.6e-280;
 Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 MAGSLTGLLLQAVSWASGARPCPKSFSGYSSVVCVNCATYCDSPPTPALGTFSTRYE 80
 DB 1 MAGSLTGLLLQAVSWASGARPCPKSFSGYSSVVCVNCATYCDSPPTPALGTFSTRYE 80
 QY 81 STRSGRRMELSMGPIQANHGTGTLTLOPEQKFKVKGFGGAMTAAALNIALSPPAQ 140
 DB 61 STRSGRRMELSMGPIQANHGTGTLTLOPEQKFKVKGFGGAMTAAALNIALSPPAQ 120
 QY 141 NLLKSYSEEGIGYNIIRVPMASCDFSIRTYTYADTDDFQHLNFSLPEEDTKLTIPLI 200
 DB 121 NLLKSYSEEGIGYNIIRVPMASCDFSIRTYTYADTDDFQHLNFSLPEEDTKLTIPLI 180
 QY 201 HRAALQALQAPVSLASPTPTWLTNGAVNGKSLGQPCDIYHQTWARYFVKFLDAYA 260
 DB 181 HRAALQALQAPVSLASPTPTWLTNGAVNGKSLGQPCDIYHQTWARYFVKFLDAYA 240
 QY 261 EHKLOFWAATVNAENEPSAGLLSGYPCQCLGTFPEHORDFIARDLGPTLANSTHNNVRLML 320
 DB 241 EHKLOFWAATVNAENEPSAGLLSGYPCQCLGTFPEHORDFIARDLGPTLANSTHNNVRLML 300
 QY 321 DDORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATIGETHRLFPNTMLFASAA 380
 DB 301 DDORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATIGETHRLFPNTMLFASAA 360
 QY 381 CVGSKFWEQSVRLGSDRGMOYSHSIITNLLYHVVGWTDNMLNPEGGPNVNRVDS 440
 DB 361 CVGSKFWEQSVRLGSDRGMOYSHSIITNLLYHVVGWTDNMLNPEGGPNVNRVDS 420
 QY 441 IIVDTKDTFYKQPMFYHLGHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSAAVVV 500
 DB 421 IIVDTKDTFYKQPMFYHLGHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSAAVVV 480
 QY 501 NRSSKDVPLTIKDPVAGFLETISPGYSIHTYLWRRQ 536
 DB 481 NRSSKDVPLTIKDPVAGFLETISPGYSIHTYLWRRQ 516

RESULT 6
 AAY99599
 ID AAY99599 standard; Protein; 516 AA.
 XX
 AC AAY99599;
 XX
 DT 08-SEP-2000 (first entry)
 XX

Human lysosomal glycoprotein glucocerebrosidase (GCS).

Human; lysosomal glycoprotein glucocerebrosidase; Gaucher's disease; GCS; beta-D-glucosyl-N-acylsphingosine glycohydrolase; neuroprotective; hepatotropic; gene therapy.

Homo sapiens.

Key Location/Qualifiers
 Peptide 1..19
 Protein /label= Signal_peptide
 20..516
 /label= Mature_protein

US6074864-A.

13-JUN-2000.

13-JAN-1994; 94US-0186256.

23-DEC-1987; 87US-0137796.
 05-FEB-1990; 90US-0474307.
 06-AUG-1992; 92US-0925333.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Gins EI, Maysak KA, Eliason WK, Lamarca ME, Martin B;
 WPI; 2000-430396/37.
 N-PSDB; AAA48444.

New glycosylated human glucocerebrosidase useful for treating Gaucher's disease, is produced by transfecting Spodoptera frugiperda cells with baculovirus containing cDNA encoding glucocerebrosidase -

Claim 4; Fig 1; 12pp; English.

The present sequence is the human lysosomal glycoprotein glucocerebrosidase (GCS, beta-D-glucosyl-N-acylsphingosine glycohydrolase). Mutation or deficiency of the GCS gene leads to Gaucher's disease. In the present invention, a Spodoptera frugiperda insect cell infected with a recombinant baculovirus vector containing a the coding sequence for the present protein has been developed. This cell line is useful for large-scale production of GCS, which could be used to treat Gaucher's disease via gene therapy.

Sequence 516 AA;

Query Match 96.4%; Score 2762; DB 21; Length 516;
 Best Local Similarity 100.0%; Pred. No. 4.6e-280;
 Matches 516; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 MAGSLGILLQLQAVSWASGAPCPKSPGYSSVVCVNCNATYCDSPDPFPALGTFRSYE 80
 Db 1 MAGSLGILLQLQAVSWASGAPCPKSPGYSSVVCVNCNATYCDSPDPFPALGTFRSYE 60

Qy 81 STRSGRMELSMGPQOANTGTGLLTLPQPKFQKVGFGGAMTDAALNTLALSPPAQ 140
 Db 61 STRSGRMELSMGPQOANTGTGLLTLPQPKFQKVGFGGAMTDAALNTLALSPPAQ 120

Qy 141 NLLKSYFEEGIGYNIIRVPWASGDFSIPTYTYADPPDQFQHNFSLPEDTKLKIPLI 200
 Db 121 NLLKSYFEEGIGYNIIRVPWASGDFSIPTYTYADPPDQFQHNFSLPEDTKLKIPLI 180

Qy 201 HRALQIAQRPVSLASPWTPTLTKTNGAVNGKSLGQPGDIYHOTWARYPVKFLDAYA 260
 Db 181 HRALQIAQRPVSLASPWTPTLTKTNGAVNGKSLGQPGDIYHOTWARYPVKFLDAYA 240

Qy 261 EHLQFWATANEPSAGLSGYPQCLGTPPEHQDFIARDLGPFLANSTHHNVRLLML 320
 Db 241 EHLQFWATANEPSAGLSGYPQCLGTPPEHQDFIARDLGPFLANSTHHNVRLLML 300

Qy 321 DQORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 380

Db 301 DQORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATLGETHRLFPNTMLFASEA 360
 Qy 381 CVGSKFWDQSVRLGSDWRGMQYSHSITNLLYHYVVGWTDNMLALNPEGGPNWYRNFVDS 440
 Db 361 CVGSKFWDQSVRLGSDWRGMQYSHSITNLLYHYVVGWTDNMLALNPEGGPNWYRNFVDS 420
 Qy 441 IIVDITKDTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQKNDLDAVALMHPDGSVAVVYL 500
 Db 421 IIVDITKDTFYKQPMFYHLGHFSKFIPEGSQRVGLVASQKNDLDAVALMHPDGSVAVVYL 480
 Qy 501 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLWRRQ 536
 Db 481 NRSSKDVPLTIKDPVAGFLETISPYSIHTYLWRRQ 516

RESULT 7

AAG78410
 ID AAG78410 standard; Protein; 516 AA.
 AC AAG78410;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Amino acid sequence of human glucocerebrosidase.
 KW Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
 KW Farber disease; Tay-Sachs disease; Niemann-Pick disease;
 KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
 KW Gaucher disease; glycosylation.

XX Homo sapiens.
 OS
 PN WO200177307-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-US11144.
 XX
 PR 06-APR-2000; 2000US-195598P.
 XX
 PA (CYTO-) CYTOCLONAL PHARM INC.
 XX
 PI Berent SL;
 XX
 DR WPI; 2002-041252/05.
 DR N-PSDB; AAH99924.

Novel expression system useful for producing glucocerebrosidase for treating Gaucher's disease, comprises an insect cell transformed with a vector encoding glucocerebrosidase that synthesizes glucocerebrosidase

Example 1; Page 65-67; 74pp; English.

This invention relates to an expression system comprising an insect cell transformed with a vector encoding glucocerebrosidase (GC) that synthesizes clinically effective GC. Recombinant GC is useful for treating individuals with deficiencies in GC by introducing the clinically effective form of recombinant GC from insect cells into individuals. The expression system is stably transformed and provides a consistently higher level of expression of GC than in baculovirus or mammalian cell expression systems, and proper glycosylation modifications for GC, requiring no enzymatic carbohydrate remodeling to be clinically effective. The expression system is more effective, economical and simple for the manufacture of recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack of lysosomal enzymes contributes to some of the many genetically inherited lysosomal storage diseases, such as Fabry disease, Farber disease, Tay-Sachs disease, Niemann-Pick disease, Hunter syndrome, Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease. This amino acid sequence represents that of human glucocerebrosidase. See also AAG78411 for the amino acid sequence of human

RESULT 9
 AAG78411
 ID AAG78411 standard; Protein; 516 AA.
 XX AAG78411;
 XX 13-FEB-2002 (first entry)
 XX Amino acid sequence of human glucocerebrosidase.
 DE Human; glucocerebrosidase; lysosomal enzyme; alpha-galactosidase;
 KW Farber disease; Tay-Sachs disease; Niemann-Pick disease;
 KW Schindler disease; Hunter syndrome; Sly syndrome; Hurler syndrome;
 KW Gaucher disease; glycosylation.
 XX Homo sapiens.
 OS
 PN WO200177307-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-US111144.
 XX
 PR 06-APR-2000; 2000US-195598P.
 XX
 PA (CYTO-) CYTOCONAL PHARM INC.
 XX
 PI Berent SL;
 XX
 DR WPI; 2002-041292/05.
 DR N-PSDB; AAH99925.
 XX
 FT Novel expression system useful for producing glucocerebrosidase for
 FT treating Gaucher's disease, comprises an insect cell transformed with a
 FT vector encoding glucocerebrosidase that synthesizes glucocerebrosidase
 FT
 XX
 PS Example 1; Page 70-72; 74pp; English.
 XX
 CC This invention relates to an expression system comprising an insect
 CC cell transformed with a vector encoding glucocerebrosidase (GC) that
 CC synthesizes clinically effective GC. Recombinant GC is useful for
 CC treating individuals with deficiencies in GC, by introducing the
 CC clinically effective form of recombinant GC from insect cells into
 CC individuals. The expression system is stably transformed and
 CC provides a consistently higher level of expression of GC than in
 CC baculovirus or mammalian cell expression systems, and proper
 CC glycosylation modifications for GC, requiring no enzymatic
 CC carbohydrate remodeling to be clinically effective. The expression
 CC system is more effective, economical and simple for the manufacture of
 CC recombinant GC. Glucocerebrosidase is a lysosomal enzyme and a lack
 CC of lysosomal enzymes contributes to some of the many genetically
 CC inherited lysosomal storage diseases, such as Fabry disease, Farber
 CC disease, Tay-Sachs disease, Niemann-Pick disease, Hunter syndrome,
 CC Schindler disease, Sly syndrome, Hurler syndrome and Gaucher disease.
 CC This amino acid sequence represents that of human
 CC glucocerebrosidase. See also AAG78410 for the amino acid sequence of
 CC human glucocerebrosidase, noting the difference at residue 514. This
 CC sequence has histidine at this position, and AAG78410 has arginine.
 XX
 SQ Sequence 516 AA:
 Query Match 96.28; Score 2757; DB 23; Length 516;
 Best Local Similarity 99.88; Pred No 1.6e-279;
 Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 21 MAGSLTGLLLQAVSWAGAPCPKPGYSSVVCNATYCDSDPPTFFALGTFSSYE 80
 Db 1 MAGSLTGLLLQAVSWAGAPCPKPGYSSVVCNATYCDSDPPTFFALGTFSSYE 60
 QY 81 STRSGRRMELSMGPQIANTHTGTLTLPQEQKFKVKFGGAMTDAALNLTALSPPAQ 140
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 61 STRSGRRMELSMGPQIANTHTGTLTLPQEQKFKVKFGGAMTDAALNLTALSPPAQ 120
 QY 141 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDDFQLHNFSLPEEDTKLKIPLI 200
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTPDDFQLHNFSLPEEDTKLKIPLI 180
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 201 HRALQLAQRVPVSLIASPWTPTWLKINGAVNGKSLKGQPGDIYHOTWARYFYKFLDAYA 260
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 HRALQLAQRVPVSLIASPWTPTWLKINGAVNGKSLKGQPGDIYHOTWARYFYKFLDAYA 240
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 261 EHKLQFNAVTAENEPSAGLLSGYFQCLGFTPEHORDFIARDLGPTLANTHNVRLML 320
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 EHKLQFNAVTAENEPSAGLLSGYFQCLGFTPEHORDFIARDLGPTLANTHNVRLML 300
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 321 DQRLLLPHWAKVVLTDPEAAKVHGIHVHVLDFLAPAKATLGETHRLFPNTMLPASEA 380
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 DQRLLLPHWAKVVLTDPEAAKVHGIHVHVLDFLAPAKATLGETHRLFPNTMLPASEA 360
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 381 CVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDNMLNPEGGNWRNFDSP 440
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 CVGSKFWEQSVRLGSDWRGMQYSHSIITNLLYHVVGWTDNMLNPEGGNWRNFDSP 420
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 441 IIVDTKDTFYKQPMFYHLGHFSKFPIEGSORVGLVASQKNDLDAVALMHPDGSVAVVYL 500
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 IIVDTKDTFYKQPMFYHLGHFSKFPIEGSORVGLVASQKNDLDAVALMHPDGSVAVVYL 480
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 501 NRSSKDVPLTIKDPVAVGFLETISPGYSIHTYLVRRQ 536
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 481 NRSSKDVPLTIKDPVAVGFLETISPGYSIHTYLVRRQ 516
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 RESULT 10
 AAP81987
 ID AAP81987 standard; protein; 516 AA.
 XX AAP81987;
 XX
 DT 17-DEC-2001 (updated)
 DT 12-OCT-1990 (first entry)
 XX
 DE Recombinant human lysosomal glucocerebrosidase.
 XX
 KW lysosomal glycoprotein glucocerebrosidase ; Gaucher's disease; ss.
 OS synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 489..489 /label-old_seq(Pro)
 FT Misc-difference 514..514 /label-old_seq(His)
 XX
 XX USN7137796-N.
 PD 02-AUG-1988.
 XX
 PF 23-DEC-1987; 87US-0192026.
 PR 23-DEC-1987; 87US-0137796.
 XX
 PA (USSH) US DEPT HEALTH AND HUMAN SERVICES.
 XX
 FI Glans EI;
 DR WPI; 1988-264274/37.
 DR N-PSDB; AAN80113.
 XX
 PT Recombinant glucocerebrosidase -
 PT useful for treating Gaucher's disease
 PS Disclosure; ; 2pp; English.
 XX
 CC The protein produced by the baculovirus expression system lacks

CC carbohydrate moieties which would be added in a eucaryotic host.
 CC The prokaryotic system produces glucocerebrosidase in far higher
 CC amt than e.g. COS cells. Recombinant protein prodn also avoids
 CC the need to purify from placenta.
 CC See also AAN80112.
 CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 XX
 XX Sequence 516 AA:
 SQ

QY 289 GFTPEHORDFIARDGPTLANSTHNHVLMLDDORLLPHAKVVLTDPEAAKYVHGIA 348
 DB 345 GFTPEHORDFIARDGPTLANSTHNHVLMLDDORLLPHAKVVLTDPEAAKYVHGIA 404
 QY 349 VHYLDFLAPAKATLGTHRLFPNTMLFASACVGSKFWEQSVRLGSDWDRGMQYSHSIIT 408
 DB 405 VHYLDFLAPAKATLGTHRLFPNTMLFASACVGSKFWEQSVRLGSDWDRGMQYSHSIIT 464
 QY 409 NLLYHVVGWTDNMLALNPEGGPNVRFVDSPIIVDITKDTFYKQPMFVHLGHFSKFIPE 468
 DB 465 NLLYHVVGWTDNMLALNPEGGPNVRFVDSPIIVDITKDTFYKQPMFVHLGHFSKFIPE 524
 QY 469 GSQRVGLVASQKNDLDAVALMHPDGSAAVVVNLNRSKDVPLTIKDPVAGFLETISPGYSI 528
 DB 525 GSQRVGLVASQKNDLDAVALMHPDGSAAVVVNLNRSKDVPLTIKDPVAGFLETISPGYSI 584
 QY 529 HTYLWRRQ 536
 DB 585 HTYLWRRQ 592

RESULT 12
 AAU07885
 ID AAU07885 standard; protein: 498 AA.
 XX AC AAU07885;
 XX 28-JAN-1997 (first entry)
 XX DE Remodelled, recombinant glucocerebrosidase (r-GCR).
 XX KW Remodelling; glycosylation; glucocerebrosidase; GCR; anaemia;
 KW lysosome deficiency; glycolipid hydrolysis; Gaucher's disease;
 KW splenomegaly; hepatomegaly; skeletal disorder; thrombocytopenia;
 KW placenta-derived; targetting.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Misc-difference 495
 FT /label= substitution
 FT /note= "His replaces Arg, found at this posn. in
 naturally occurring placental GCR (p-GCR)"
 XX US5549892-A.
 XX 27-AUG-1996.
 XX PF 23-DEC-1988; 88US-0289589.
 XX 21-JUN-1993; 93US-0080855.
 XX 23-DEC-1988; 88US-0289589.
 XX 22-DEC-1989; 89US-0455507.
 XX 21-AUG-1991; 91US-0748283.
 XX (GENZ) GENZYME CORP.
 XX PI Friedman B, Hayes M;
 XX WPI; 1996-401555/40.
 XX Re-modelled glucocerebrosidase has increased fucose and N-acetyl
 glucosamine residues - effectively alleviates clinical symptoms of
 Gaucher's disease at a lower dose than naturally occurring enzyme
 XX Example 1; Column 9-12; 9pp; English.
 XX AAU07885 is a recombinant version of the commercially available
 remodelled glucocerebrosidase, p-GCR (derived from placental tissue).
 CC The recombinant glucocerebrosidase (r-GCR) has a His-Arg substitution
 CC at posn. 495 and is remodelled to increase the level of glycosylation
 CC of the enzyme, having four complex carbohydrate chains and increased

CC fucose and N-acetyl glucosamine (Glc-Nac) residues compared to p-GCR.
 CC The remodelled r-GCR has improved pharmacokinetics, including
 CC approximately a two-fold increase in affinity for its target cells
 CC (Kupfer cells). Remodelled r-GCR is useful in treating patients with
 CC Gaucher's disease.

XX Sequence 498 AA;

Query Match 93.0%; Score 2665; DB 17; Length 498;
 Best Local Similarity 99.8%; Pred. No. 6.4e-270;
 Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 40 ARCPKPSFGYSVVCVNCATYCDSDFPPTPALGTFSRYSTRSRMRMELSMGPIQANH 99
 DB 1 ARCPKPSFGYSVVCVNCATYCDSDSDFPPTPALGTFSRYSTRSRMRMELSMGPIQANH 60
 QY 100 TGTGLLTQLQPEQKFKVKGFGGAMTDAALNIALSPPAQNLLKSYFSESGIGYNIIR 159
 DB 61 TGTGLLTQLQPEQKFKVKGFGGAMTDAALNIALSPPAQNLLKSYFSESGIGYNIIR 120
 QY 160 VPMASCDFSIRTYTADTDDPDLHNFSLPEEDTKLPIHLHAIQALQORPVSLASPT 219
 DB 121 VPMASCDFSIRTYTADTDDPDLHNFSLPEEDTKLPIHLHAIQALQORPVSLASPT 180
 QY 220 SPWLKTNAGVNGKSLKGQPGDIYHOTWARYFYKFLDAYAEHKLQFWATVAENEPSAGL 279
 DB 181 SPWLKTNAGVNGKSLKGQPGDIYHOTWARYFYKFLDAYAEHKLQFWATVAENEPSAGL 240
 QY 280 LSGYFQCLGFTPEHORDFIARDLGTLANSTHNHVLMLDDORLLPHAKVVLTDPE 339
 DB 241 LSGYFQCLGFTPEHORDFIARDLGTLANSTHNHVLMLDDORLLPHAKVVLTDPE 300
 QY 340 AAKYVHGIAVHWYLDPLAPAKATLGTHRLFPNTMLFASACVGSKFWEQSVRLGSDWDRG 399
 DB 301 AAKYVHGIAVHWYLDPLAPAKATLGTHRLFPNTMLFASACVGSKFWEQSVRLGSDWDRG 360
 QY 400 MOYSHSIITNLLYHVVGWTDNMLALNPEGGPNVRFVDSPIIVDITKDTFYKQPMFVHL 459
 DB 361 MOYSHSIITNLLYHVVGWTDNMLALNPEGGPNVRFVDSPIIVDITKDTFYKQPMFVHL 420
 QY 460 GHFSKFTEPGSORVGLVASQKNDLDAVALMHPDGSAAVVVNLNRSKDVPLTIKDPVAGFL 519
 DB 421 GHFSKFTEPGSORVGLVASQKNDLDAVALMHPDGSAAVVVNLNRSKDVPLTIKDPVAGFL 480
 QY 520 ETISPGYSIHTYLWRRQ 536
 DB 481 ETISPGYSIHTYLWRRQ 497

RESULT 13
 AAU05696
 ID AAU05696 standard; Protein: 497 AA.
 XX AC AAU05696;

XX 24-OCT-2001 (first entry)

XX Human Mature glucocerebrosidase, GCB.

XX Human; glucocerebrosidase; GCB; lysosomal storage disease;
 KW Gaucher's disease; Fabry's disease; Farber's disease;
 KW G_m1 gangliosidosis; Tay-Sachs disease; Niemann-Pick disease;
 KW Shindler disease; Hunter syndrome; Sly syndrome; Huler syndrome;
 XX Schete syndrome.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Modified-site 19
 FT /note= "N-glycosylated"
 FT Modified-site 59
 FT /note= "N-glycosylated"
 FT Modified-site 146

FT Misc-difference 259 /note= "N-glycosylated"
 FT /note= "Encoded by TTA"
 FT Modified-site 270
 FT /note= "N-glycosylated"
 FT Misc-difference 495
 FT /label= OTHER
 FT /note= "OTHER= Arg or His"
 XX W0200149830-A2.
 PN 12-JUL-2001.
 PD 29-DEC-2000; 2000WO-DK00743.
 XX 30-DEC-1999; 99DK-0001891.
 PR 02-JUN-2000; 2000DK-0000865.
 PR 02-JUN-2000; 2000DK-0000866.
 PR 30-JUN-2000; 2000DK-0001027.
 XX (MAXY-) MAXYGEN APS.
 PA Okkels JS, Jensen AD, Halkier T, Jensen RB, Schambye HT;
 PI WPI: 2001-465259/50.
 DR N-PSDB; AAS10863.
 XX Improved lysosomal enzymes and lysosomal enzyme activators useful for
 PT treating Gaucher's disease -
 PS Claim 20; Page 96; 97pp; English.
 XX The sequence is the mature human lysosomal enzyme glucocerebrosidase,
 CC GCB. GCB is the enzyme involved in Gaucher's disease, a lysosomal
 CC storage disease. The invention relates to introducing new glycosylation
 CC sites into lysosomal enzymes/activators like GCB to improve their
 CC bioactivity. The novel polypeptides are used for the prevention and
 CC treatment of Gaucher's disease, Fabry's disease, Farber's disease,
 CC G_{M1} gangliosidosis, Tay-Sachs's disease, Niemann-Pick disease, Shindler
 CC disease, Hunter syndrome, Sly syndrome, Hurler and Huler/Scheie
 CC syndromes.
 XX Sequence 497 AA;
 SQ
 Query Match 93.0%; Score 2664; DB 22; Length 497;
 Best Local Similarity 99.8%; Pred. No. 8.2e-270;
 Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 40 ARPCIPKSGYSSVVCNATYCSFDPPTPALCTFSRYESTSRGRMELSMGPIQANH 99
 DB 1 ARPCIPKSGYSSVVCNATYCSFDPPTPALCTFSRYESTSRGRMELSMGPIQANH 60
 QY 100 TGTGGLLLTLPQEQKQKVKYKGGAMTDAALNLLALSPPAQNLLKSYFSEEGIGYNIIR 159
 DB 61 TGTGGLLLTLPQEQKQKVKYKGGAMTDAALNLLALSPPAQNLLKSYFSEEGIGYNIIR 120
 QY 160 VPMASCDPSIRTYTADTPDFDLHNFSLPREDTKLKIPLTHRALQLAQRVSVLLASPTW 219
 DB 121 VPMASCDPSIRTYTADTPDFDLHNFSLPREDTKLKIPLTHRALQLAQRVSVLLASPTW 180
 QY 220 SPTLTKTNVANGKSLGKQGDYHQWARYFVKFLDAYAEHLQFVAVTAENPSAGL 279
 DB 181 SPTLTKTNVANGKSLGKQGDYHQWARYFVKFLDAYAEHLQFVAVTAENPSAGL 240
 QY 280 LSGYFQCLGFTPEHQRDIFARDLGPTLANSTHNRVLLMLDDQRLLLPHWAKVYLTDPE 339
 DB 241 LSGYFQCLGFTPEHQRDIFARDLGPTLANSTHNRVLLMLDDQRLLLPHWAKVYLTDPE 300
 QY 340 AAKYVHGIAVHWYLDLPAKATIGETHRLFPNTMLFASEACVSGKFEQSVRLGSDWRG 399
 DB 301 AAKYVHGIAVHWYLDLPAKATIGETHRLFPNTMLFASEACVSGKFEQSVRLGSDWRG 360
 QY 400 MQYSHSIITNLLYHVGVGTDNLLALNPGPNVNRVFDSPILVDITKDTFYKQPMFYHL 459

Db 361 MQYSHSIITNLLYHVGVGTDNLLALNPGPNVNRVFDSPILVDITKDTFYKQPMFYHL 420
 QY 460 GHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSVAVVVLNRSSKDVPPLTIKDPVAGFL 519
 Db 421 GHFSKFIPEGSRQVGLVASQKNDLDAVALMHPDGSVAVVVLNRSSKDVPPLTIKDPVAGFL 480
 QY 520 ETISPGYSIHTYLWRRQ 536
 Db 481 ETISPGYSIHTYLWRRQ 497
 RESULT 14
 AAUB6020
 ID AAU86020 standard; Protein; 497 AA.
 XX AAU86020;
 AC
 XX 21-MAY-2002 (first entry)
 DT
 XX Human glucocerebrosidase.
 DE
 XX Human; glucocerebrosidase; GCB; antidiabetic; cytostatic; cancer;
 KW antiinflammatory; dermatological; immunosuppressive; antirheumatic;
 KW antiarthritic; antithyroid; detergent; textile; food product; vaccine;
 KW agrochemical; feed product; viral pathogen; parasite; autoantigen;
 KW autoimmune disease; insulin-dependent diabetes mellitus; gene therapy;
 KW autoimmune thyroid disorder; inflammatory condition; sperm antigen.
 XX Homo sapiens.
 OS
 XX WO200202597-A2.
 PN
 XX 10-JAN-2002.
 PD
 XX 29-JUN-2001; 2001WO-DK00459.
 PF
 XX 30-JUN-2000; 2000DK-0001027.
 PR 14-JUL-2000; 2000DK-0001092.
 PR 29-DEC-2000; 2000WO-DK00743.
 PR 09-FEB-2001; 2001WO-DK00090.
 XX (MAXY-) MAXYGEN APS.
 PA (MAXY-) MAXYGEN HOLDINGS LTD.
 PA
 XX Okkels JS, Jensen AD, Van Den Hazel B;
 PI WPI: 2002-195745/25.
 DR N-PSDE; ABK40235.
 DR
 XX Novel glycosylated polypeptides having improved properties as compared
 PT to polypeptide of interest e.g. improved shelf-life, stability, and
 PT reduced renal clearance, comprise an additional glycosylation site -
 XX
 XX Example 2; Page 90-92; 136pp; English.
 PS
 XX The invention relates to novel glycosylated polypeptides (I) into which
 CC an additional glycosylation site (GS) has been introduced. (I) comprises
 CC the primary structure: NH₂-X-Pp-COOH or NH₂-Px-X-Py-COOH where:
 CC X = a peptide addition comprising or contributing to a GS; Pp = a
 CC polypeptide; Px = an N-terminal part of Pp; and Py = a C-terminal of Pp.
 CC Also described is a method of improving one or more selected properties
 CC of Pp, by preparing a nucleotide sequence encoding a polypeptide with the
 CC primary structure NH₂-X-Pp-COOH, where X is a peptide addition comprising
 CC or contributing to a GS that is capable of conferring the selected
 CC improved properties to Pp, expressing the nucleotide sequence in a
 CC suitable host cell, optionally conjugating the polypeptide to a
 CC second non-peptide moiety and recovering the polypeptide. The
 CC glycosylated polypeptides have improved properties compared to the
 CC polypeptide of interest. Pp is chosen from an antibody or its fragment,
 CC plasma protein, erythrocyte or thrombocyte protein, cytokine, growth
 CC factor, profibrinolytic protein, protease inhibitor, antigen, enzyme,
 CC ligand, receptor, or a hormone, or is a microbial enzyme chosen from

CC protease, amylase, amyloglucosidase, pectinase, lipase and cutinase. Pp
 CC can be a therapeutic polypeptide useful in human or veterinary therapy, a
 CC diagnostic polypeptide used in diagnostics, an industrial polypeptide
 CC useful in the manufacture of products such as detergents, household
 CC articles, textile, food products, agrochemicals, feed products,
 CC polypeptides or proteins associated with viral pathogens, antigens from
 CC parasites, autoantigens associated with autoimmune diseases such as
 CC insulin-dependent diabetes mellitus; autoimmune thyroid disorders;
 CC inflammatory conditions; sperm antigens; and bacterial and cancer
 CC antigens. When the polypeptide is an in vivo glycosylated
 CC polypeptide which does not comprise any other type of non-peptide
 CC moiety, a nucleotide sequence encoding the polypeptide can be used in
 CC gene therapy. When the polypeptide Pp is an antigen, the polypeptide
 CC can be provided in the form of a vaccine. AAU86020-AAU86128 represent
 CC the peptide addition sequences, and related protein sequences of
 CC the invention.
 XX
 XX Sequence 497 AA;

Query Match 93.08; Score 2664; DB 23; Length 497;
 Best Local Similarity 99.68; Pred. No. 8.2e-270;
 Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 40 ARCPKPSFGYSSVVCVNCATYCDSPDPTPALGTFSYSTRSGRMELSMGPIQANH 99
 DB 1 ARCPKPSFGYSSVVCVNCATYCDSPDPTPALGTFSYSTRSGRMELSMGPIQANH 60
 QY 100 TGTGLLLTLQPEOKFKVKYGGCGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 159
 DB 61 TGTGLLLTLQPEOKFKVKYGGCGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 120
 QY 160 VPMASCDFSIRTYTADPTDFDLHNFSLPEEDTKLKIPLIHRLAQLAQRVSLASPT 219
 DB 121 VPMASCDFSIRTYTADPTDFDLHNFSLPEEDTKLKIPLIHRLAQLAQRVSLASPT 180
 QY 220 SPTWLKTNAGVNGKSLGQPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 279
 DB 181 SPTWLKTNAGVNGKSLGQPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGL 240
 QY 280 LSGYPCQCLGFTPEHQRDIARDLGPTLANSTHNNVRLMLDDQRLPHAKVVLTDPE 339
 DB 241 LSGYPCQCLGFTPEHQRDIARDLGPTLANSTHNNVRLMLDDQRLPHAKVVLTDPE 300
 QY 340 AAKYVHGIAVHYLDLFLAPAKATIGETHRLFPNTMLFASEACVGSKFWEQSVRLGSDWRG 399
 DB 301 AAKYVHGIAVHYLDLFLAPAKATIGETHRLFPNTMLFASEACVGSKFWEQSVRLGSDWRG 360
 QY 400 MOYSHSIITNLLYHVVGWTDNMLALNPEGGPNVNRNFDSPDIIVDITKDTFYKQPMFYHL 459
 DB 361 MOYSHSIITNLLYHVVGWTDNMLALNPEGGPNVNRNFDSPDIIVDITKDTFYKQPMFYHL 420
 QY 460 GHFSKFIPEGSQVRGLVASQKNDLDAVALMHPDGSAAVVVVLNRSSKDVPLTIKDPAYGFL 519
 DB 421 GHFSKFIPEGSQVRGLVASQKNDLDAVALMHPDGSAAVVVVLNRSSKDVPLTIKDPAYGFL 480
 QY 520 ETISPGYSIHTYLWRRQ 536
 DB 481 ETISPGYSIHTYLWRRQ 497

RESULT 15
 AAU05668
 ID AAU05668 standard; Protein: 497 AA.

XX
 XX AC AAU05668;
 XX
 XX DT 24-OCT-2001 (first entry)
 XX
 XX DE Human glucocerebrosidase, GCB, mutant K194N.
 XX
 XX KW Human; glucocerebrosidase; GCB; lysosomal storage disease;
 KW Gaucher's disease; Fabry's disease; Farber's disease;
 KW G_m1 gangliosidosis; Tay-Sach's disease; Niemann-Pick disease;

KW Shindler disease; Hunter syndrome; Sly syndrome; Huler syndrome;
 KW Scheie syndrome; mutant; mutein; K194N.

OS Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 19 /note= "N-glycosylated"
 FT Modified-site 59 /note= "N-glycosylated"
 FT Modified-site 146 /note= "N-glycosylated"
 FT Modified-site 194 /note= "N-glycosylated"
 FT Misc-difference 194 /note= "Wild-type Lys changed to Asn"
 FT Modified-site 270 /note= "N-glycosylated"
 FT Misc-difference 495 /label= OTHER
 FT /note= "OTHER= Arg or His"

XX WO200149830-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-DK00743.

XX 30-DEC-1999; 99DK-0001891.

XX 02-JUN-2000; 2000DK-0000865.

XX 30-JUN-2000; 2000DK-0000865.

XX 30-JUN-2000; 2000DK-0001027.

XX (MAXY-) MAXYGEN APS.

XX Okkels JS, Jensen AD, Halkier T, Jensen RB, Schambye HT;

XX WPI; 2001-465259/50.

XX Improved lysosomal enzymes and lysosomal enzyme activators useful for
 XX treating Gaucher's disease -

XX Claim 20; Page - ; 97pp; English.

XX The sequence is a mutant of human lysosomal enzyme
 CC glucocerebrosidase, GCB, which has an introduced glycosylation site.
 CC GCB is the enzyme involved in Gaucher's disease, a lysosomal
 CC storage disease. The invention relates to introducing new glycosylation
 CC sites into lysosomal enzymes/activators like GCB to improve their
 CC bioactivity. The novel polypeptides are used for the prevention and
 CC treatment of Gaucher's disease, Fabry's disease, Farber's disease,
 CC G_m1 gangliosidosis, Tay-Sach's disease, Niemann-Pick disease, Shindler
 CC disease, Hunter syndrome, Sly syndrome, Hurler and Huler/Scheie
 CC syndromes.

CC Note: The present sequence is not displayed in the specification
 CC but was created from the GCB sequence shown on page 96.

XX Sequence 497 AA;

Query Match 92.88; Score 2659; DB 22; Length 497;
 Best Local Similarity 99.68; Pred. No. 2.7e-269;
 Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 ARCPKPSFGYSSVVCVNCATYCDSPDPTPALGTFSYSTRSGRMELSMGPIQANH 99
 DB 1 ARCPKPSFGYSSVVCVNCATYCDSPDPTPALGTFSYSTRSGRMELSMGPIQANH 60
 QY 100 TGTGLLLTLQPEOKFKVKYGGCGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 159
 DB 61 TGTGLLLTLQPEOKFKVKYGGCGAMTDAALNIALSPPAQNLLKSYFSEEGIGYNIIR 120
 QY 160 VPMASCDFSIRTYTADPTDFDLHNFSLPEEDTKLKIPLIHRLAQLAQRVSLASPT 219

Db 121 VPMASCOFSIRTYADTDDFQDHNFSLEEDTKLKIPLIHRADQLAORPVSLASPT 180
QY 220 SPTWLKTNGAVNGSKGQPGDIYHQTWARYFVKFELDAYAEHKLQFWAVTAENEPSAGL 279
Db 181 SPTWLKTNGAVNGSKGQPGDIYHQTWARYFVKFELDAYAEHKLQFWAVTAENEPSAGL 240
QY 280 LSGYPFQCLGFTPEHORDFIARDLGPTLANSTHHNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 241 LSGYPFQCLGFTPEHORDFIARDLGPTLANSTHHNVRLMLDDQRLLLPHWAKVVLTDPE 300
QY 340 AAKYVHGIAVHWYLDPLAPAKATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDRG 399
Db 301 AAKYVHGIAVHWYLDPLAPAKATLGETHRLFPNTMLFASEACVSGKFWQSVRLGSDRG 360
QY 400 MOYSHSIITNLLYHVVGWTDWNLALNPEGGNWVRNFVDSPIIIVDITKDTFYKQPMFYHL 459
Db 361 MOYSHSIITNLLYHVVGWTDWNLALNPEGGNWVRNFVDSPIIIVDITKDTFYKQPMFYHL 420
QY 460 GHFSKEIPEGSORVGLVASQKNDLDAVALMHPDGSAAVVVVVLRSSKDVPLTIKDPVAGFL 519
Db 421 GHFSKEIPEGSORVGLVASQKNDLDAVALMHPDGSAAVVVVVLRSSKDVPLTIKDPVAGFL 480
QY 520 ETISPGYSIHTYLWRRQ 536
Db 481 ETISPGYSIHTYLWRRQ 497

Search completed: February 20, 2003, 11:18:10
Job time : 41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: February 20, 2003, 11:19:22 ; Search time 14 Seconds
(without alignments)
978.157 Million cell updates/sec

Title: US-10-024-197-25
Perfect score: 2865
Sequence: 1 MEFSSPSECPKPLRSVSI.....GFLETISPGSYHTLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pap.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2675	93.4	592	10	US-09-753-126-4
2	2664	93.0	497	10	US-09-753-126-1
3	92	3.2	399	10	US-09-815-242-13600
4	91.5	3.2	427	9	US-10-062-523-16
5	89	3.1	740	9	US-10-051-909-37
6	88.5	3.1	4303	9	US-09-904-968A-2
7	88	3.1	556	10	US-09-771-161A-245
8	87.5	3.1	551	10	US-09-897-214-8
9	86.5	3.0	600	10	US-09-801-368-158
10	85.5	3.0	731	9	US-09-738-626-4854
11	85	3.0	615	10	US-09-815-242-13747
12	84.5	2.9	667	9	US-09-759-056-2
13	84.5	2.9	667	9	US-09-901-812-2
14	84.5	2.9	667	9	US-10-227-884-80
15	84.5	2.9	728	9	US-10-106-092-4
16	84	2.9	1233	9	US-09-738-626-4312
17	83	2.9	341	9	US-09-738-626-6673
18	83	2.9	791	9	US-09-908-193-18
19	82.5	2.9	658	9	US-09-759-056-5

20	82.5	2.9	658	9	US-09-901-812-5	Sequence 5, Appli
21	81.5	2.8	953	9	US-10-118-984-43	Sequence 43, Appl
22	81.5	2.8	953	10	US-09-728-721-43	Sequence 43, Appl
23	81.5	2.8	2802	9	US-09-808-602-81	Sequence 81, Appl
24	80.5	2.8	375	9	US-09-738-626-6846	Sequence 6846, Ap
25	80.5	2.8	803	9	US-09-801-220-2	Sequence 2, Appli
26	80.5	2.8	828	9	US-09-738-626-5038	Sequence 5038, Ap
27	80.5	2.8	1137	9	US-10-174-590-518	Sequence 518, App
28	80.5	2.8	1137	9	US-10-176-758-518	Sequence 518, App
29	80.5	2.8	1137	9	US-10-173-706-518	Sequence 518, App
30	80.5	2.8	1137	9	US-10-175-737-518	Sequence 518, App
31	80.5	2.8	1137	9	US-10-175-738-518	Sequence 518, App
32	80.5	2.8	1137	9	US-10-175-752-518	Sequence 518, App
33	80.5	2.8	1137	9	US-10-176-482-518	Sequence 518, App
34	80.5	2.8	1137	9	US-10-176-757-518	Sequence 518, App
35	80.5	2.8	1137	9	US-10-176-913-518	Sequence 518, App
36	80.5	2.8	1137	9	US-10-180-552-518	Sequence 518, App
37	80.5	2.8	1137	9	US-10-180-557-518	Sequence 518, App
38	80.5	2.8	1137	9	US-10-173-700-518	Sequence 518, App
39	80.5	2.8	1137	9	US-10-174-579-518	Sequence 518, App
40	80.5	2.8	1137	9	US-10-174-582-518	Sequence 518, App
41	80.5	2.8	1137	9	US-10-174-588-518	Sequence 518, App
42	80.5	2.8	1137	9	US-10-175-739-518	Sequence 518, App
43	80.5	2.8	1137	9	US-10-175-740-518	Sequence 518, App
44	80.5	2.8	1137	9	US-10-175-740-518	Sequence 518, App
45	80.5	2.8	1137	9	US-10-175-743-518	Sequence 518, App

ALIGNMENTS

RESULT 1
US-09-753-126-4
; Sequence 4, Application US/09753126
; Patent No. US20020127219A1
; GENERAL INFORMATION:
; APPLICANT: OKKELS, JENS SIGURD
; APPLICANT: JENSEN, ANNE DAM
; APPLICANT: HALKIER, TORBEN
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME
; TITLE OF INVENTION: ACTIVATORS
; FILE REFERENCE: 31-00060005
; CURRENT APPLICATION NUMBER: US/09753,126
; FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PA 1999 01891
; PRIOR FILING DATE: 1999-12-30
; PRIOR APPLICATION NUMBER: 60/174,652
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PA 200 00865
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/210,984
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 60/211,124
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: PA 2000 01027
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/217,497
; PRIOR FILING DATE: 2000-07-11
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: SapC-linker-GCB polypeptide
US-09-753-126-4

Query Match 93.4%; Score 2675; DB 10;
Best Local Similarity 92.3%; Pred. No. 6.4e-244;
Matches 506; Conservative 8; Mismatches 14; Indels 20; Gaps 2;


```

RESULT 4
US10-062-523-16
; Sequence 16, Application US/10062523
; Publication No. US2003002327A1
; GENERAL INFORMATION:
;
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Dendritic Enriched Secreted Lymphocyte Activation
; TITLE OF INVENTION: Molecule
; FILE REFERENCE: EP448P2
; CURRENT APPLICATION NUMBER: US/10/062,523
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/267,523
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: PCT/US00/21130
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/190,062
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 09/369,248
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: PCT/US99/02415
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/244,110
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 60/078,572
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/073,962
; PRIOR FILING DATE: 1998-02-06

```

```

1  RESULT 5
2  'US/10-051-909-37
3  ; Sequence 37, Application US/10051909
4  ; Publication No. US20020199217A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Allen, Steve
7  ; APPLICANT: Helentjaris, Tim
8  ; APPLICANT: Hitz, Bill
9  ; APPLICANT: Kinney, Tony
10 ; APPLICANT: Tingey, Scott
11 ; TITLE OF INVENTION: Plant Sugar Transport Proteins
12 ; FILE REFERENCE: Bb1163 US CIP
13 ; CURRENT APPLICATION NUMBER: US/10/051,909
14 ; CURRENT FILING DATE: 2002-01-17
15 ; PRIOR APPLICATION NUMBER: 60/083,044
16 ; PRIOR FILING DATE: April 24, 1998
17 ; NUMBER OF SEQ ID NOS: 38
18 ; SOFTWARE: Microsoft office 97
19 ; SEQ ID NO 37
20 ; LENGTH: 740
21 ; TYPE: PRT
22 ;

```

•

RESULT 8
US-09-897-214-8
; Sequence 8, Application US/09897214
; Patent No. US2002007679A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Edward C.
; APPLICANT: Pressnell, Scott R.
; TITLE OF INVENTION: Leucine-Rich Repeat Proteins, Zlrr7,
; FILE REFERENCE: 01-27 Zlrr8, and Zlrr9
; CURRENT APPLICATION NUMBER: US/09/897,214
; CURRENT FILING DATE: 2001-07-02
; PRIOR FILING DATE: 60/215,446
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-214-8

Query Match 3.1%; Score 87.5; DB 10; Length 551;
Best Local Similarity 20.8%; Pred. No. 5.4;
Matches 115; Conservative 44; Mismatches 171; Indels 223; Gaps 27;
QY 28 LLLQAVSWASGARPCIPKSGYS-SVVCVNA-----TYCDSDP 67
DB 6 LLLLSAGAACPLPCVCONSESLTCAHRGLLFPVPPNDRRTVELKADNFQALGP 65
QY 68 PTPPALG-----TFSEYSTRGRMELSMGPIQANHTGGLTLLTLPQEQKQYKGG 122
DB 66 PDFRNMGLVDLTLSRNATTRIGARAFGLDLESLSLHGNRLVELGT-----G 114
QY 123 AMTDAALNIALS-----PPAONLLKSYFSEEGIGYNIIR-VPMSCDFSIPTY 173
DB 115 SLRGPVNLQHLISGNLQRIAPGAFDDFLES-LEDLDLSYNNLRQVPWAGIGAMPALHT 173
QY 174 Y-----ADTPDF-----QLHNFSPPEE--DTKLKLIHRALOAQRPVSLASPT 219
DB 174 LNLHNLIDALPGFAOLGQLSRDLTNSRLATLPDPLFSRGRDAEASAPLVLSFG 233
QY 220 SP-----TWLKTNGAVNGSLKGOPDIYHQTWARYVFKLDAYAEHKLQFWATAE 272
DB 234 NPLHCNCELLWRLRLRPDDLETCAAPG-----LAGRY-----FWAV-PE 273
QY 273 NEPSAGLLSGYPQCIQFTPEHORFTARDLGPTLANSTHNVRLMLDDOELLPHWAK 332
DB 274 GE-----FSC-----EPPLAR-----HTQRLWLEGQRTLRCRA- 304
QY 333 VLTDPAAKYVGHVAVHWYLDLAPAKATLGETH--RLFPNTMLF----- 376
DB 305 --LGDPA-----TMHW-----VGPDDRLVGNSSRARAPPNGTLEIGVTGAGDAGGYTC 351
QY 377 -----ASEACV-----GSKFWQSYR- 392
DB 352 IATNPAGATARVELRVLALPHGNGSSAEGGRFGPSDIAASARTAAEGGTTLESEPAQV 411
QY 393 -----LGSDRG-----MOYSHSIITNLLY-----HVGWTDWN 421
DB 412 TEVTATSGLVSGPGKPADPVMMFQIQYNSSEDETLLIYRIVPASSHHLLKHLVPGADYD 471
QY 422 ---LALNPEGGPN 431
DB 472 LCLLALSPAAGFS 484

RESULT 9
US-09-801-368-158
; Sequence 158, Application US/09801368
; Patent No. US20020128250A1

; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 158
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-158

Query Match 3.0%; Score 86.5; DB 10; Length 600;
Best Local Similarity 19.5%; Pred. No. 7.7;
Matches 37; Conservative 26; Mismatches 62; Indels 65; Gaps 8;
QY 306 TLANSTHNVRLMLDDOELLPHWAK-----KALPAQWIRCDVRKDFRVLGKFSVVIADPA----- 350
DB 309 TLGCSAHCIK-----TLG-ETHRLFPNTMLFASACVSKFW--EQSVRLG----- 394
QY 351 WYLDLAPAKA-----TLG-ETHRLFPNTMLFASACVSKFW--EQSVRLG----- 394
DB 351 WNIHMLPYCTCNDIELLGLPLHQLDSEGIIF-----LWVTGRAELGKESLNNG 401
QY 395 -----SWDRGMQYSHSIITNLLYHVVGWTDNLALNPEGGPNVNFVDSPIIVDTK 447
DB 402 YNVINEYSWIKTNQLGRTIVTGTGHNLSKEHLLVGLKGNFKWINKHIDVLIYSMTIR 461
QY 448 DTFYKQPMFY 457
DB 462 ETSRKPDELY 471

RESULT 10
US-09-738-626-4854
; Sequence 4854, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 4854
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4854

Query Match 3.0%; Score 85.5; DB 9; Length 731;
Best Local Similarity 19.0%; Pred. No. 13;
Matches 77; Conservative 49; Mismatches 118; Indels 161; Gaps 23;
QY 154 GYNITRVPMASCDSEIRYTYADTDDPOLHNFSPPEEDTKLK---IPLIHRAQLAQRP 210
DB 164 GWNASOHPMRSGGSLWELFTPGIEEGEVYKFAVOTREGQRKADPMARRA-ELAPAT 222
QY 211 VSLLAS---PMTSPWLKTNAGVNGKSLGQPGDIYH-----QWARYF----- 252
DB 223 GSIVASSEYQWQDSEWRLERSQTD---LASKPMSVYEVHLGSRWGNKYNEDLATELVDY 278
QY 253 -----VKFIDAAEHKLFQWAVTAENEPSAGLLSGYPCQICG-----TPEHOR 296
DB 279 VADLGTYHVEFLPV-AEH-----PFGG---SWGQVGTGYAYAPTSRWGTPDQFR 322
QY 297 ----DPIARDLGPTL-----ANSTHNVRLMLDDORLL-LPHWAKVLTDD----- 337
DB 323 ALVDAPFHARGIVMDVWPAHFPPKDDWALARDFDGEALYHPDWRGEGKDWGTLVDFDGR 382
QY 338 PEAAKYVHGIYVHW-----YLD-----FL 356
DB 383 NEVRNVLNVALYWEIEPHIDGLRDVAVASMLYLDYSREHGEWEPNIYGGRENLEAVQFL 442
QY 357 APAKATLGTETHRLFPNTMLFASEACVSKFWQSGVRLGSDWRGMOYSHSIITNLLYHVVG 416
DB 443 QEMNATV---LRLHPGALTAIEE---STSW-PGVTAPTWDDGLGFS----- 481
QY 417 WTDWNLALNPEGGPNVNRNFVDSPIIVDTTKDTFYKQPM---FYH 458
DB 482 -LKNWM-----GWMD-----TLEYFSKNPVHRAFFH 507

RESULT 11
US-09-815-242-13747
; Sequence 13747, Application US/09815242
; Patent No. US20070061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13747
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13747

Query Match 3.0%; Score 85; DB 10; Length 615;
Best Local Similarity 27.5%; Pred. No. 11;
Matches 39; Conservative 16; Mismatches 67; Indels 20; Gaps 6;
QY 279 LLSGYPFCQGLGTEPHORDFIARDLGTPLANSTHNVRLMLDDORLL-----LPHWA 331
DB 92 LANGVAACLNSTQSRQQL-----EVMAGCTGQIRLLYIAPERLMLDNFLDLAHWN 145
QY 332 KVLITDPEAAKYVHGIYVHNYLDFLAPAKATLGTETHRLFPNTMLFASEACVSKFWQSV 391
DB 146 PVLLAVDEA-----HCIS-QWGHDF-RPEYAAAGQLQRFPALPFMALATATADDTTQDII 199
QY 392 R-IGSWDRGMQYSHSIITNLLY 412
DB 200 RLLGLNDPLTIQISSFDRPNIRY 221

RESULT 12
US-09-759-056-2
; Sequence 2, Application US/09759056
; Patent No. US20020156252A1
; GENERAL INFORMATION:
; APPLICANT: Pennica, Diane
; APPLICANT: Smith, Victoria
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: No. US20020156252A1el STRA6 Polypeptides
; FILE REFERENCE: GENENT.2827A2
; CURRENT APPLICATION NUMBER: US/09/759,056
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/197089
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/175849
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: 60/228914
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-759-056-2

Query Match 2.9%; Score 84.5; DB 9; Length 667;
Best Local Similarity 19.0%; Pred. No. 14;
Matches 117; Conservative 65; Mismatches 202; Indels 233; Gaps 31;
QY 4 SSPSRE---ECPK-----PLSRVYSIMAGSLTGLLLIQAQVSWAS-GAR-- 41
DB 131 SAPSDGKTEAPRGANKILGLFYAALYPLAACCA-TAGHTAAHLLGLSTLSWAHLGVQVW 189
QY 42 -----PCIPKSGYSSVVCVNCATYCDSDPPTPALGTFSTRYESTSRGRMELSMGPIQ 96
DB 190 QRAECPOVPKIKYKYSLIA-----SLPLLLGLGLSLWY-----PVQ 226
QY 97 -----ANHTGTGLLLTLQPEQKFKQKVGFGAMTDAALANILALSPPAONLLKSYSEE 151
DB 227 LVRSFSRRTGAG-----SKGLQSSYSEYLRNLL-----CRKLLGSSYHTSK 268
QY 152 GIGYNIIRVPMASCDSEIRYTYADTDDPOLHNFSPPEEDTKLKFIHRAQLAQRPV 211
DB 269 HGFLSWARVCLRHCIY-----TPQP-----GFLP-----LKLVLSTLTLGTATYQV 310

```

QY 212 SLASPTSTWLTGTNGAVN-----GKSLKGQPGD-----IYHOTWA----- 249
DB 311 ALLLLGVVPTIQKVRAGVTTDVSYLLAGFIVLSEDKQEVVELVKHLLWALEVCYISAL 370
QY 250 -----RYFKFLDAYAEHK-----LQFWATNAENPSAGLL-----SGY--PF 285
DB 371 VLSCLLTFVLMSLVTHRTNLRALHRAALDLSPLHRSPPHSPRQAIFCWMFSFAYQTAF 430
QY 286 QCLGFTPEHORFIARD-----LQPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 339
DB 431 ICLGLLVQOIIFFLGTALAFVLMFVL-----HGRNLLFRSLESSWPFWLTAL----- 481
QY 340 AAKVYHGIADVHTLDFLAPAKATIGETHRLFPNT-----MLFAEACVGSKEWE 388
DB 482 -AVILQNMAAHW--FL-----ETHDGHQPQLTNRRLVLYAATFLLFPLNLVGA----- 526
QY 389 QSVRLGSWDRGMOYSHSITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDITKD 448
DB 527 ---MVATW-----RVLLSALYNAIHLGQMDLSLLPPRA-----ATLD 560
QY 449 TFYKQPMFYHLGHFSKFIPEGSQVRGLVASOKNDLDAVLMHPDGSVVVVLNRSKQVP 508
DB 561 ----PGYTYRNFLKI-----EVSQSHPAMTAFCSLLLOAQSLLP 596
QY 509 LTIKDPVAVGFELETISPG 525
DB 597 RTMAAQ-----DSLRLPG 609

```

RESULT 13

```

US-09-901-812-2
; Sequence 2, Application US/09901812
; Patent No. US20020173461A1
; GENERAL INFORMATION:
; APPLICANT: Pennica, Diane
; APPLICANT: Polakis, Paul
; APPLICANT: Szeto, Wayne
; TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO
; FILE REFERENCE: GENENT 083A
; CURRENT APPLICATION NUMBER: US/09/901,812
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/228,914
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 09/759,056
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/175,849
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: 60/197,089
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-901-812-2

```

Query Match 2.98; Score 84.5; DB 9; Length 667;

Best Local Similarity 19.08; Pred No. 14; Matches 117; Conservative 65; Mismatches 202; Indels 233; Gaps 31;

```

QY 4 SPSRE---ECPK-----PLSRVSTMAAGSLTGLLLQAVSNAS-GAR-- 41
DB 131 SAPSQDKTEAPRGAWKILGFYAAIYPLAAC- TAGHTAAHLGLSTLSWAHLGVQVW 189
QY 42 -----PCIPKSGYSSVVCNATYCDSPDPTFPALGTFYSRSTRSGRRMELSMGPIQ 96
DB 190 QRAECPQVPKIKYKYSLLA-----SLPLLGLGLFLSLW-----PVQ 226
QY 97 -----ANHTGTGLLTLTQPEOKFOVKFGFGAMTDAAALNIALSPPAQNLLKSYSEE 151

```

```

DB 227 LVRSFSRRTGAG-----SKGLSSYSEEVLRNLL-----CRKKLGSSVHTSK 268
QY 152 GIGYNIIRVPMASCDSEIRTYTYADTDDFQLHNFSLPEDTKLPIHLRALQLAORPV 211
DB 269 HGFLSNARVCLRHCIY-----TPDP-----GFHLP-----LKLVLSATLTGTAIYQV 310
QY 212 SLASPTSTWLTGTNGAVN-----GKSLKGQPGD-----IYHOTWA----- 249
DB 311 ALLLLGVVPTIQKVRAGVTTDVSYLLAGFIVLSEDKQEVVELVKHLLWALEVCYISAL 370
QY 250 -----RYFKFLDAYAEHK-----LQFWATNAENPSAGLL-----SGY--PF 285
DB 371 VLSCLLTFVLMSLVTHRTNLRALHRAALDLSPLHRSPPHSPRQAIFCWMFSFAYQTAF 430
QY 286 QCLGFTPEHORFIARD-----LQPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 339
DB 431 ICLGLLVQOIIFFLGTALAFVLMFVL-----HGRNLLFRSLESSWPFWLTAL----- 481
QY 340 AAKVYHGIADVHTLDFLAPAKATIGETHRLFPNT-----MLFAEACVGSKEWE 388
DB 482 -AVILQNMAAHW--FL-----ETHDGHQPQLTNRRLVLYAATFLLFPLNLVGA----- 526
QY 389 QSVRLGSWDRGMOYSHSITNLLYHVVGWTDWNLALNPEGGPNWVRNFVDSPIIVDITKD 448
DB 527 ---MVATW-----RVLLSALYNAIHLGQMDLSLLPPRA-----ATLD 560
QY 449 TFYKQPMFYHLGHFSKFIPEGSQVRGLVASOKNDLDAVLMHPDGSVVVVLNRSKQVP 508
DB 561 ----PGYTYRNFLKI-----EVSQSHPAMTAFCSLLLOAQSLLP 596
QY 509 LTIKDPVAVGFELETISPG 525
DB 597 RTMAAQ-----DSLRLPG 609

```

RESULT 14

```

US-10-227-884-80
; Sequence 80, Application US/10227884
; Publication No. US20030027988A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerrietsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3530PIC79
; CURRENT APPLICATION NUMBER: US/10/227,884
; CURRENT FILING DATE: 2002-08-25
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656

```

[illegible][illegible]


```
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 60/146963
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/149320
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149638
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/151733
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/164418
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/166361
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/169445
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169835

Query Match      2.9%  Score 84.5; DB 9; Length 667;
Best Local Similarity 19.0%  Pred. No. 14;
Matches 117; Conservative 65; Mismatches 202; Indels 233; Gaps 31;

QY 4 SSPSRE---ECPK-----PLSRYSIMAGSLTGLLLLOAYSHAS-GAR---41
Db 131 SAPSQDKTEAPGAKWILGLFYAALYPLAAC-A-TAGTAHLLGSLTSLWAHLGVQVW 189
QY 42 -----PCIPKSGYSSVVCVNCATYCDSDPPTPALGTSTRYESTRSGRMELSMGPIQ 96
Db 190 QRAECQVQPIKYYSLLA-----SLPLLGLGLSLWY-----PVQ 226
QY 97 -----ANHTGTGLLLTLOPEQKQKVKFGGAMTDAALNIALSPPAQNLLKSYFSEE 151
Db 227 LVRSFSRTGAG-----SKLOSSYSEELRNLL-----CRKLGSGSYHTSK 268
QY 152 GIGYNIIRVPMASCDPSIRTYADTPDDFOLHNFSLPEEDTKLIPLIHRAQLAQRPV 211
Db 269 HGFLSWARVCLRHCIY-----TPQ-----GFHLP-----LKLVSATLTGTAIYQV 310
QY 212 SLASAPWTSPTWTKYNGAVN-----GKSLKGQPGD-----IVHOTWA-----249
Db 311 ALLLVGVVPTIQKVRAGVTTDVSYLLAGFIVLSDEQEVVELVKHLLWALEVCYISAL 370
QY 250 -----RYFKFEDAYAEHK-----LOFWAYTAENEPSAGLL-----SGY---PF 285
Db 371 VLSCLLTLFLVLMRSLVTHRTNLRALHRCALDLSPHRSPPHSRQAFQWMSFSAIQTA 430
QY 286 QCLGFTPEHQRFIARD-----LGPTLANSTHNVRLMLDDQRLLLPHWAKVVLTDPE 339
Db 431 ICLGLLVQIIFLGTALAFVILMPVL-----HGRNLLFRSLESSWPFWLTLAL-----481
QY 340 AAKYVHGIHVHVLDFLAPAKATLGETHRLFPNT-----MLFASEACVGSKEWE 388
Db 482 -AVILONMAAHWV--FL-----ETHDGHQPTNRRVLYAATFLLFLPLNLVGA-----526
QY 389 QSVRLGSDRGMOYSHSIITNLLYHVYVGTWDLNALNPEGPNWVRNFVDSPIIVDITKD 448
Db 527 ---MVATW-----RVLLSALYNATHLGOMDLSLLPPRA-----ATLD 560
QY 449 TFYKQPMFYHLGHFSKFIPESQSVGLVASKNDLDAVALMHHPDGSVVVVVNLNRSSKDPV 508
Db 561 -----PGYTYRNFLKI-----EVSQSHPEAMTAFCSLLQAOQLLP 596
QY 509 LTIKDPVAGVLETISPG 525
Db 597 RTWAATQ-----DSLRF 609
```

RESULT 15

```
US-10-106-092-4
; Sequence 4, Application US/10106092
; Patent No. US20020156044A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Tufaro, Frank
; TITLE OF INVENTION: USE OF EXT GENES FOR THE TREATMENT OF
; CANCER AND OTHER DISEASES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/106,092
; FILING DATE: 25-Mar-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 920041.414C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-106-092-4

Query Match      2.9%  Score 84.5; DB 9; Length 728;
Best Local Similarity 20.8%  Pred. No. 16;
Matches 65; Conservative 42; Mismatches 97; Indels 109; Gaps 16;

QY 243 IYHOTWARYFVKELDAYAEHKLQFW--AVTAENE-----PSAG 278
Db 24 IYIT--LFSIVLGLIATGMEQFWPHSIESNDWNVKEKSIKRDVPVRLPADSPIPERG 81
QY 279 LLSYPPQCL-----GFTPEHQ-RDFI-----ARDLGPTLAN--STHHNVRLMLDDQ 324
Db 82 DLSCRMHTCFDYVRCGFNPKIKYIYALKKYVDDEGVSVSNTISREYNELIMATSDS- 140
QY 325 LLLPHWAKVVLTOPEAAKYVHGIHVHVLDFLAPAKATLGETHRLFPNTMLFASACVGS 384
Db 141 -----DYTTDDINRACLFVPSIDVNLQNTLRKETKETA---171
QY 385 KFWQSQRVLSWDRGMOYSHSIITNLLYHVY--GWTDLNLAN-PE-----GPNW- 432
Db 172 ---QAMAQLSRWDGRTNH-----LLENMLPGGPPDYNTALDVPDRALLAGGGSFTW 221
QY 433 -----VRNEVDSPIIVDITTKTFYKQPMFYHLGHFSKFIPESQSVGLVASKNDLDAV 486
Db 222 YRQGYDYSIPVYSPLSAEVDLPEKGFGRQYFL-----LSSQVGLUPEYREDLEAL 272
QY 487 ALMHPDGSATVVV 499
Db 273 QVKH--GESVLVL 283

Search completed: February 20, 2003, 11:24:18
Job time : 19 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:08:31 ; Search time 14 Seconds
(without alignments)
1587.951 Million cell updates/sec

Title: US-10-024-197-25

Perfect score: 2865
Sequence: 1 MEFSPPSRECPKPLSRVSI.....GFLTSPGSIHTYLWRRQ 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2860	99.8	536	1	P04062 homo sapien
2	2416.5	84.3	515	1	GLCM_MOUSE
3	123.5	4.3	562	1	EXG2 YEAST
4	105	3.7	2700	1	ZAN2 HUMAN
5	104.5	3.6	1738	1	CO4_MOUSE
6	99.5	3.5	466	1	GUN5_TREPU
7	99	3.5	668	1	GALC_MOUSE
8	97.5	3.4	669	1	GALC_MACMU
9	97	3.4	641	1	IND_ARTGO
10	96	3.4	1711	1	PTPO_RAT
11	95	3.3	512	1	HMZ2_ARATH
12	95	3.3	514	1	C11A_DASAM
13	94.5	3.3	669	1	GALC_HUMAN
14	94.5	3.3	822	1	PHK_IACIA
15	92.5	3.2	6486	1	TYCC_BACBR
16	92	3.2	774	1	FECA_ECOLI
17	91	3.2	1774	1	MSAS_PENPA
18	89.5	3.1	1748	1	POLR_ELV
19	89.5	3.1	2025	1	TTC3_HUMAN
20	89	3.1	555	1	HUTU_XANCP
21	88.5	3.1	350	1	DCUP_SYNY3
22	88.5	3.1	464	1	YRG5_CAEEL
23	88.5	3.1	4303	1	PKD1_HUMAN
24	88	3.1	556	1	PDPR_HUMAN
25	88	3.1	565	1	SAPA_HAHIN
26	87.5	3.1	874	1	STLM_SCHPO
27	87.5	3.1	920	1	AKH1_MAIZE
28	87	3.0	391	1	PCL_ECTHA
29	87	3.0	627	1	HUTU_XANAC
30	87	3.0	555	1	CACP_COLLI
31	87	3.0	1274	1	SRF3_BACSU
32	86.5	3.0	600	1	SPO8_YEAST
33	86.5	3.0	1448	1	CPSA_ARATH

GLCM_HUMAN	STANDARD;	PRT;	536 AA.
AC	P04062;		
DT	01-NOV-1986 (Rel. 03, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Glucosylceramidase precursor (EC 3.2.1.45) (Beta-glucocerebrosidase)		
DE	(Acid beta-glucosidase) (D-glucosyl-N-acylsphingosine glucosylase)		
DE	(Alglucerase) (Imiglucerase).		
GN	GBA OR GC.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Placenta;		
RX	MEDLINE=86042651; PubMed=3864160;		
RA	Sorge J., West C., Westwood B., Beutler E.;		
RT	"Molecular cloning and nucleotide sequence of human		
RL	glucocerebrosidase cDNA.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 82:7289-7293(1985).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Hepatoma;		
RX	MEDLINE=86085859; PubMed=3001061;		
RA	Tsuji S., Choudary P.V., Martin B.M., Winfield S., Barranger J.A.,		
RT	Guinn E.I.;		
RL	"Nucleotide sequence of cDNA containing the complete coding sequence		
RL	for human lysosomal glucocerebrosidase.";		
RL	J. Biol. Chem. 261:50-53(1986).		
RN	[3]		
RP	SEQUENCE OF 1-11 FROM N.A.		
RX	MEDLINE=88195776; PubMed=3359914;		
RA	Reiner O., Wigderson M., Horowitz M.;		
RT	"Structural analysis of the human glucocerebrosidase genes.";		
RL	DNA 7:107-116(1988).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	Horowitz M.;		
RL	Submitted (MAY-1988) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	SEQUENCE OF 1-45 FROM N.A.		
RX	MEDLINE=88074307; PubMed=3687939;		
RA	Sorge J.A., West C., Kuhl W., Treger L., Beutler E.;		
RT	"The human glucocerebrosidase gene has two functional ATG initiator		
RL	codons.";		
RL	Am. J. Hum. Genet. 41:1016-1024(1987).		
RN	[6]		
RP	SEQUENCE OF 40-44.		
RC	TISSUE=Placenta;		
RA	Martin B.M., Murray G.J., Colligan J.E., Raum M., Brady R.O.,		
RT	Barranger J.A.;		
RL	"Structural studies of human placental glucocerebrosidase.";		
RL	Fed. Proc. 43:1869-1869(1984).		

34	86	3.0	553	1	HT2A_HUMAN	Q13049 homo sapien
35	86	3.0	1286	1	AIDA_ECOLI	Q03155 escherichia
36	86	3.0	1365	1	GTFS_STRDO	P29336 streptococc
37	85	3.0	563	1	BZNB_PSEFL	P51853 pseudomonas
38	85	3.0	608	1	RECQ_SALTU	P40724 salmonella
39	85	3.0	643	1	RHOP_MOUSE	Q61085 mus musculu
40	85	3.0	689	1	UVRA_PSELE	P52087 pseudomonas
41	85	3.0	820	1	CHIA_ALTSO	P32823 alteromonas
42	84.5	2.9	456	1	ENO_MYCPN	P75189 mycoplasma
43	84.5	2.9	589	1	SPY_DROME	O44783 drosophila
44	84.5	2.9	718	1	EXT2_HUMAN	Q93063 homo sapien
45	84.5	2.9	863	1	PHSG_MYCTU	Q10639 mycobacteri

ALIGNMENTS

RT [7] SEQUENCE OF 469-520.
RL TISSUE-Placenta;
RX MEDLINE-86149363; PubMed-3456607;
RA Dinur T., Osiecki K.M., Legler G., Gatt S., Desnick R.J.,
RA Grabowski G.A.;
RT "Human acid beta-glucosidase: isolation and amino acid sequence of a
RT peptide containing the catalytic site";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1660-1664(1986).
RN [8]
RP REVIEW ON GD VARIANTS.
RX MEDLINE-94163182; PubMed-8118460;
RA Horowitz M., Zimran A.;
RT "Mutations causing Gaucher disease";
RL Hum. Mutat. 3:1-11(1994).
RN [9]
RP REVIEW ON GD VARIANTS.
RX MEDLINE-97044507; PubMed-8889578;
RA Beutler E., Gelbart T.;
RT "Glucocerebrosidase (Gaucher disease)";
RL Hum. Mutat. 8:207-213(1996).
RN [10]
RP REVIEW ON GD VARIANTS.
RX MEDLINE-99459154; PubMed-10527671;
RA Tayebi N., Stone D.L., Sidransky E.;
RT "Type 2 Gaucher disease: an expanding phenotype";
RL Mol. Genet. Metab. 68:209-219(1999).
RN [11]
RP REVIEW ON GD VARIANTS.
RX MEDLINE-20115339; PubMed-10649495;
RA Stone D.L., Tayebi N., Orvisky E., Stubblefield B., Madike V.,
RA Sidransky E.;
RT "Glucocerebrosidase gene mutations in patients with type 2 Gaucher
RT disease";
RL Hum. Mutat. 15:181-188(2000).
RN [12]
RP VARIANTS GD TYR-255.
RX MEDLINE-90343254; PubMed-1974409;
RA Beutler E., Gelbart T.;
RT "Gaucher disease associated with a unique KpnI restriction site:
RT identification of the amino-acid substitution";
RL Ann. Hum. Genet. 54:149-153(1990).
RN [13]
RP VARIANTS GD.
RX MEDLINE-90274868; PubMed-1972019;
RA Hong C.M., Ohashi T., Yu X.J., Weiler S., Barranger J.A.;
RT "Sequence of two alleles responsible for Gaucher disease";
RL DNA Cell Biol. 9:233-241(1990).
RN [14]
RP VARIANTS GD.
RX MEDLINE-93162655; PubMed-8432337;
RA Beutler E., Gelbart T., West C.;
RT "Identification of six new Gaucher disease mutations";
RL Genomics 15:203-205(1993).
RN [15]
RP VARIANTS GD HIS-535.
RX MEDLINE-94379086; PubMed-7916532;
RA Choy F.Y.M., Wei C., Applegarth D.A., McGillivray B.C.;
RT "DNA analysis of an uncommon missense mutation in a Gaucher disease
RT patient of Jewish-Polish-Russian descent";
RL Am. J. Med. Genet. 51:156-160(1994).
RN [16]
RP VARIANTS GD ASN-438.
RX MEDLINE-94156356; PubMed-8112750;
RA Beutler E., Gelbart T.;
RT "Two new Gaucher disease mutations";
RL Hum. Genet. 93:209-210(1994).
RN [17]
RP VARIANTS GD SER-409 AND CYS-457.
RX MEDLINE-94357595; PubMed-8076951;
RA Tuteja R., Tuteja N., Lilliu F., Bambi B., Galanello R., Cao A.,
RA Baralle F.E.;
RT "Y418C: a novel mutation in exon 9 of the glucocerebrosidase gene of

RT a patient with Gaucher disease creates a new Bgl I site";
RL Hum. Genet. 94:314-315(1994).
RN [18]
RP VARIANTS GD D-215; T-221; R-241; Q-296; C-324; G-417 AND N-419.
RX MEDLINE-96382740; PubMed-8790604;
RA Beutler E., Demina A., Gelbart T.;
RT "Glucocerebrosidase mutations in Gaucher disease";
RL Mol. Med. 1:82-90(1994).
RN [19]
RP VARIANTS GD SER-409; HIS-448; PRO-483 AND CYS-502.
RX MEDLINE-95353281; PubMed-7627184;
RA Cormand B., Vilageliu L., Burguera J.M., Balcells S.,
RA Gonzalez-Duarte R., Grinberg D., Chabas A.;
RT "Gaucher disease in Spanish patients: analysis of eight mutations";
RL Hum. Mutat. 5:303-309(1995).
RN [20]
RP VARIANTS GD SER-217.
RX MEDLINE-95353289; PubMed-7627192;
RA Choy F.Y.M., Wei C.;
RT "Identification of a new mutation (P178S) in an African-American
RT patient with type 2 Gaucher disease";
RL Hum. Mutat. 5:345-347(1995).
RN [21]
RP VARIANTS GD SER-409; LEU-426; LEU-433 AND PRO-483.
RX MEDLINE-97092100; PubMed-8937765;
RA Morar B., Lane A.B.;
RT "The molecular characterization of Gaucher disease in South Africa";
RL Clin. Genet. 50:78-84(1996).
RN [22]
RP VARIANTS GD LEU-54; GLU-85 AND SER-227.
RX MEDLINE-96271984; PubMed-8829654;
RA Kim J.-W., Liou B.B., Lai M.-Y., Ponce E., Grabowski G.A.;
RT "Gaucher disease: identification of three new mutations in the Korean
RT and Chinese (Taiwanese) populations";
RL Hum. Mutat. 7:214-218(1996).
RN [23]
RP VARIANTS GD HIS-352 AND GLN-398.
RX MEDLINE-96271993; PubMed-8829663;
RA Cormand B., Vilageliu L., Balcells S., Gonzalez-Duarte R., Chabas A.,
RA Grinberg D.;
RT "Two novel (1098InsA and Y313H) and one rare (R359Q) mutations
RT detected in exon 8 of the beta-glucocerebrosidase gene in Gaucher's
RT disease patients";
RL Hum. Mutat. 7:272-274(1996).
RN [24]
RP VARIANTS GD THR-435.
RX MEDLINE-97044520; PubMed-8889591;
RA Amaral O., Pinto E., Fortuna M., Lacerda L., Sa Miranda M.C.;
RT "Type 1 Gaucher disease: identification of N396T and prevalence of
RT glucocerebrosidase mutations in the Portuguese";
RL Hum. Mutat. 8:280-281(1996).
RN [25]
RP VARIANTS GD LEU-437 AND ILE-530.
RX MEDLINE-96373808; PubMed-8780099;
RA Seeman P.J.V., Finckh U., Hoepfner J., Lakner V., Liebisch I.,
RA Grau G., Rolfs A.;
RT "Two new missense mutations in a non-Jewish Caucasian family with type
RT 3 Gaucher disease";
RL Neurology 46:1102-1107(1996).
RN [26]
RP VARIANTS GD LEU-414 AND THR-441.
RX MEDLINE-97325776; PubMed-9182788;
RA Cormand B., Grinberg D., Gort L., Fiumara A., Barone R., Vilageliu L.,
RA Chabas A.;
RT "Two new mild homozygous mutations in Gaucher disease patients:
RT clinical signs and biochemical analyses";
RL Am. J. Med. Genet. 70:437-443(1997).
RN [27]
RP VARIANTS GD V-76; E-85; W-87; W-159; S-227; I-252 AND P-483.
RX MEDLINE-97360189; PubMed-9217217;
RA Choy F.Y.M., Humphries M.L., Shi H.;
RT "Identification of two novel and four uncommon missense mutations
RT among Chinese Gaucher disease patients";

Query Match 99.8%; Score 2860; DB 1; Length 536;
 Best Local Similarity 99.8%; Pred. No. 2.7e-233;
 Matches 535; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEFSRSPRECKPLSRVSMAGSLGGLLLQAVSWAGARPCIPKSGYSSVVCNAT 60
 DB 1 MEFSRSPRECKPLSRVSMAGSLGGLLLQAVSWAGARPCIPKSGYSSVVCNAT 60
 QY 61 YCDSPDPTFPALGTFRSYESTSGRMELSMGPIQANHTGTGLLLTLOPEKQFVKGF 120
 DB 61 YCDSPDPTFPALGTFRSYESTSGRMELSMGPIQANHTGTGLLLTLOPEKQFVKGF 120
 QY 121 GCAMTDAALNIALSPAPQNLKLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTDD 180
 DB 121 GCAMTDAALNIALSPAPQNLKLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTDD 180
 QY 181 FOLHNFSLPEEDTKLTIPLHRAQLAQPVSLLASPTWTLKTNAGVNGKSLGQP 240
 DB 181 FOLHNFSLPEEDTKLTIPLHRAQLAQPVSLLASPTWTLKTNAGVNGKSLGQP 240
 QY 241 GDVYHQTWARYVFKFLDAYAEHLQFQWATYAEPSAGLLSGYPCQCLGFTPEHORDFA 300
 DB 241 GDVYHQTWARYVFKFLDAYAEHLQFQWATYAEPSAGLLSGYPCQCLGFTPEHORDFA 300
 QY 301 RDLGPTLANSTHHNVRLLMDQRLLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
 DB 301 RDLGPTLANSTHHNVRLLMDQRLLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAK 360
 QY 361 ATLGTHRLFPNTMLFASACVSKFEQSVRLGSDWDRGMOYSHSIITNLLYHVVGTDW 420
 DB 361 ATLGTHRLFPNTMLFASACVSKFEQSVRLGSDWDRGMOYSHSIITNLLYHVVGTDW 420
 QY 421 NLALNPEGGNNWVNFVDSPIIIVDITKDFYKQPMFYHLGHFSKFIPEGSQVGLVASQK 480
 DB 421 NLALNPEGGNNWVNFVDSPIIIVDITKDFYKQPMFYHLGHFSKFIPEGSQVGLVASQK 480
 QY 481 NDLDAVALMHPDGSAAVVVNLNRSSKDVPILTIDKPAVGLETISPGYSIHTYLMWRQ 536
 DB 481 NDLDAVALMHPDGSAAVVVNLNRSSKDVPILTIDKPAVGLETISPGYSIHTYLMWRQ 536

RESULT 2
 GLCM_MOUSE STANDARD; PRT; 515 AA.

AC 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Glucosylceramidase precursor (EC 3.2.1.45) (Beta-glucocerebrosidase)
 DE (Acid beta-glucosidase) (D-glucosyl-N-acylsphingosine glucosylase).
 GN GBA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89296941; PubMed=2740343;
 RA O'Neill R.R., Tokoru T., Kozak C.A., Brady R.O.;
 RT "Comparison of the chromosomal localization of murine and human
 RT glucocerebrosidase genes and of the deduced amino acid sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:5049-5053(1989).
 RN [2]
 RP SEQUENCE OF 1-18 FROM N.A.
 RC STRAIN=BALE/c;
 RX MEDLINE=92272748; PubMed=1317175;
 RA Carstea E.D., Murray G.J., O'Neill R.R.;
 RT "Molecular and functional characterization of the murine
 RT glucocerebrosidase gene.";
 RL Biochem. Biophys. Res. Commun. 184:1477-1483(1992).
 CC -!- CATALYTIC ACTIVITY: D-glucosyl-N-acylsphingosine + H(2)O = D-
 glucose + N-acylsphingosine.

CC -!- SUBCELLULAR LOCATION: LYSOSOMAL, MEMBRANE BOUND.
 CC -!- SIMILARITY: BELONGS TO FAMILY 30 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M24119; AAA37671.1; -;
 CC EMBL; M89949; AAA37665.1; -;
 CC PIR; A32931; A32931.
 CC MGD; MGI:95665; Gba.
 CC InterPro; IPR001139; GH_30.
 CC Pfam; PF02055; Glyco_hydro_30; 1.
 CC PRINTS; PR00843; GLHYDRLASE30.
 KW Hydrolase; Glycosidase; Sphingolipid metabolism; Glycoprotein;
 KW Lysosome; Membrane; Gaucher disease; Signal.
 FT SIGNAL 1 19 GLUCOSYL CERAMIDASE.
 FT CHAIN 20 515 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 515 AA; 57621 MW; 7CDD9176085FE2CB CRC64;

Query Match 84.3%; Score 2416.5; DB 1; Length 515;
 Best Local Similarity 86.4%; Pred. No. 6.1e-196;
 Matches 446; Conservative 30; Mismatches 39; Indels 1; Gaps 1;

QY 21 MAGSLTGILLQAVSWAGARPCIPKSGYSSVVCNATYCDSPDPTFPALGTFRSYE 80
 DB 1 MAARLIGFFLEQAVSWAGARPCIPKSGYSSVVCNATYCDSPDPTFPALGTFRSYE 60
 QY 81 STRSRGRMELSMGPIQANHTGTGLLLTLOPEKQFVKFGGAMTDAALNIALSPAPQ 140
 DB 61 STRSRGRMELSMGPIQANHTGTGLLLTLOPEKQFVKFGGAMTDAALNIALSPAPQ 120
 QY 141 NLLKSYFSEEGIGYNIIRVPMASCDFSIRTYTADTDDFOLHNFSLPEEDTKLTIPL 200
 DB 121 KLLRSYFSTNGIEYNIIRVPMASCDFSIRTYTADTDDFOLHNFSLPEEDTKLTIPL 180
 QY 201 HRAQLAQRPSVLLASPTWTLKTNAGVNGKSLGKQPGDIYHQTWARYVFKFLDAYA 260
 DB 181 HQALMSSRPISLASPWTSPWLKTNAGVNGKSLGKQPGDIYHQTWARYVFKFLDAYA 240
 QY 261 EHKLOFWATYAEPSAGLLSGYPCQCLGFTPEHORDFTARDLPTLANSTHHNVRLLML 320
 DB 241 KYGLRFWATYAEPSAGLLSGYPCQCLGFTPEHORDFTARDLPTLANSTHHNVRLLML 299
 QY 321 DQORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATLGTHRLFPNTMLFAS 380
 DB 300 DQORLLPHWAKVLTDPAAKYVHGIAVHWYLDLAPAKATLGTHRLFPNTMLFAS 359
 QY 381 CVGSKFEQSVRLGSDWDRGMOYSHSIITNLLYHVVGTDWNLALNPEGGNNWVNFVDS 440
 DB 360 CVGSKFEQSVRLGSDWDRGMOYSHSIITNLLYHVVGTDWNLALNPEGGNNWVNFVDS 419
 QY 441 IIVDTKDFYKQPMFYHLGHFSKFIPEGSQVGLVASOKNDLDAVALMHPDGSAAVVV 500
 DB 420 IIVDTKDFYKQPMFYHLGHFSKFIPEGSQVGLVASOKNDLDAVALMHPDGSAAVVV 479
 QY 501 NRSSKDVPILTIDKPAVGLETISPGYSIHTYLMWRQ 536
 DB 480 NRSSKDVPILTIDKPAVGLETISPGYSIHTYLMWRQ 515

RESULT 3

QY	46	KSEGYSVVVCNATYCDSFDPPTFPALGTFSGRYESTRSGRMWELSM-----GPIQANH	100
Db	4	KSFFSAFLVLCLSKSTQ-----GVGTTEKESLSP--LELNILQNKEASYANDT	52
QY	101	GT-----GLLLT-----LOPEKQFKVKFGGGAMTDA	127
Db	53	ITVKGITIGWLVPITYTPSLYRNATSLAKOONSSNSISIVDEFTCKTLTGNTSLT--	110
QY	128	AALNILALSPPAQLL--LKVSFSEGI-----CYNLIIRPVMASCDFSIITYYADTP	178
Db	111	-----LDLNFHKTWITEDDFAQKINGFNLRIPIGYWAWK-----QNTD	150
QY	179	DFQLHN--FSLPEDTKIKPLIHRALOLAQRPSVLSLASPWTSPTWLKTNGAV-----	230
Db	151	KNLYIDNIITFNDFPVSDGLQKLNNALEWAQK-----YELNVLDLHGAPGSQNGF	202
QY	231	--NGKSLKGQPQ-----DIYHOTWARYFYVKFLDAYAEHLQFMWATAEPEPSAGL-	279
Db	203	DNSGERILYDGLWLRLNNTKETLAIWRDMFOTFLN--KGDKSPVVGIIIVEPLGGKI	260
QY	280	----LSCYPFOCLGFPEHQRO-----FTARD-----LGPTLANSTHH--NVPL	317
Db	261	DVSDIETMYEAEPDLKKNQNSDNFTFIHDGFGQIGHWNLELPYQVSHHYFNLTG	320
QY	318	LMLDDORLLPHWAKVVLVDPEAAK-----YVHGIAVHWYLDFLAPAKATIGETHR	368
Db	321	ANTSSODILVDHHYEVFDQALETQAFARIENINYGDSIIKELSF---HPAVGVGWSG	377
QY	369	L-----PNTMIFASE-----ACVGSKF-----WEQSRLGSWDRGM	400
Db	378	AITDCATWLNGVGARYDGSYNTTFLTNDKPVGTCISQNSLADWTQDYR---DRVVR	433
QY	401	OYSHSIITNLHYHVCGWTOWN	421
Db	434	OFTERAQLATYSKTKTGWIIFWN	454
RESULT 4			
ID	ZAN_HUMAN	STANDARD;	PRT; 2700 AA.
AC	Q9Y493; O00218;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Zonadhesin (Fragment).		
GN	ZAN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;			
[1]			
RN	SEQUENCE OF 1-2379 FROM N.A.		
RP	MEDLINE=99018118; PubMed=9799793;		
RX	Glockner G., Scherer S., Schatttevoy R., Boright A., Weber J.,		
RA	Tsui L.C., Rosenthal A.;		
RT	"Large-scale sequencing of two regions in human chromosome 7q22:		
RT	analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci		
RT	reveals 17 genes."		
RN	Genome Res. 8:1060-1073(1998).		
RC	[2]		
RP	SEQUENCE OF 2338-2700 FROM N.A.		
RC	TISSUE=Testis;		
RX	MEDLINE=97271566; PubMed=9126492;		
RA	Gao Z., Harumi T., Garbers D.L.;		
RT	"Chromosome localization of the mouse zonadhesin gene and the human		
RT	zonadhesin gene (ZAN)."		
CC	Genomics 41:119-122(1997).		
CC	-!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA		
CC	OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR		
CC	SIGNALING.		
CC	-!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.		
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE		


```

FT CONFLICT 739 740 DL -> AI (IN REF. 5).
FT CONFLICT 838 839 P -> R (IN REF. 4).
FT CONFLICT 993 993 P -> L (IN REF. 5).
FT CONFLICT 1043 1043 D -> E (IN REF. 5).
FT CONFLICT 1119 1119 V -> A (IN REF. 7).
FT CONFLICT 1190 1190 A -> T (IN REF. 7).
FT CONFLICT 1324 1324 K -> N (IN REF. 4).
FT CONFLICT 1401 1401 R -> S (IN REF. 9).
FT CONFLICT 1442 1442 R -> K (IN REF. 4).
FT CONFLICT 1453 1453 A -> V (IN REF. 4).
SQ SEQUENCE 1738 AA; 192870 MW; D1E02AE7AB42BF6 CRC64;

Query Match 3.6%; Score 104.5; DB 1; Length 1738;
Best Local Similarity 22.6%; Pred. No. 2.6;
Matches 79; Conservative 30; Mismatches 142; Indels 99; Gaps 15;

QY 132 ILALSPAQN-----LLKSYFSEEGIGYNIIRVPM-ASG-----DFSIRTYIYADT 177
DB 23 LLLFSPVWNLGTPSLVGVQLLDAPPGQEVKGSVFLRNPKGSGSPKKDKL-----SS 76

QY 178 PDDQLHNSFSLPEEDTKLPIHLRALAQRP-VSLLA-SPWTSPTWLKTNQVANGKGS 235
DB 77 GDDFVLLSLEVPEDYR-----SCGLFLRAPHQLVQSPWLNRTAFKATETQGVNLL 131

QY 236 LKGQPGDIYQHT-----WARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPPQ 286
DB 132 FSSRRGHIFVOTDQPIYNGQVRVYFALDQKMRPSTDFLTITVNSHGLRVLKKEFT 191

QY 287 CLGTPPEHQDFIARDL-----GPTLANSTHNVRLMLDDQRLLLPHWAK 332
DB 192 S-----TSIFODAFIPDISBPGTWKISARFSDGLESNRSTHFVKVYLPNFVETKTPWKP 248

QY 333 VVLTDPE-----AARYVHGIYVHW--YLDL-----APAKATLGET 366
DB 249 YILMPSNDSIEDIOLQIARYIYKPKVQGVAYTREALDQEQKTRFLKLOAKLVEGT 308

QY 367 H-----RLFPNTMLFASACVGSKFWE-OSVRLGSM 396
DB 309 HISISKDQFOALDKINIGVRDLEGLRYAATAVIESPGGMEAEELTSM 358

RESULT 6
GUN5_THEFU
ID GUN5_THEFU STANDARD; PRT; 466 AA.
AC 001786;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase E-5 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E-4)
DE (Cellulase E-5) (Cellulase E5).
GN CELE.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxId=2021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YX;
RX MEDLINE=91258320; PubMed=1904434;
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca.";
RL J. Bacteriol. 173:3397-3407 (1991).
RN [2]
RP REVISIONS.
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 137-142 AND 157-166.
RC STRAIN=YX;
RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
RT "Activity studies of eight purified cellulases: specificity,
RT synergism, and binding domain effects.";
```

```

RL Biotechnol. Bioeng. 42:1002-1013 (1993).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L01577; AAC09379.1; -.
CC PIR: C42360; C42360.
CC HSP: P07986; IEXG.
CC InterPro: IPR001919; Bac_celose-bind.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00553; CBM_2; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 36
CC CHAIN 37 466 ENDOGLUCANASE E-5.
CC ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
CC SEQUENCE 466 AA; 49800 MW; 1CF0ADF2DEF82E CRC64;

Query Match 3.5%; Score 99.5; DB 1; Length 466;
Best Local Similarity 20.6%; Pred. No. 1.1;
Matches 97; Conservative 53; Mismatches 151; Indels 169; Gaps 30;

QY 34 VSMAGARPCIPKSFY-----SSVVCVGNATYCD-----SFDPTFPAL 73
DB 102 VSNSTTPPGGTASSGFIASGSGEPHTCTINGAPCDGSEPGGPGCTPDPGTPGT 161

QY 74 GT-PSRYESTR-SGRMELSMG-PQANHGTGTLTLOPEQKFKVKGFGGAMTAAAL 130
DB 162 GTPVERIGKVGCGTQCDHEGNPNVLRGMST-----HGQWFDHCLTD-SSL 208

QY 131 NILASPPAQNLLKSYFSEEGIGYNIIRVPMASCD-----FSIRT-----YTYAD---T 177
DB 209 DALAYDWKADIIRLSMYIQEDGYETN---PRGFTDRMHQIDMATARGLYIVDWHILT 264

QY 178 PDDQLHNSFSLPEEDTKLPIHLRALAQRPVSVLLASPTWTSPTWLKTNQVANGKSLK 237
DB 265 PGD---PHYNLDRAKTF-----AETARHAS-----KTNVLY-----EIA 297

QY 238 GQPGDIYHOTWARYFVKFLDAYAEHKLQFWAVTAENEPSAGLLSG-YPFQCLGFTPEHOR 296
DB 298 NEPNGV---SWAS-----IKSYAEVI---PVIRQRPDSVIVITRGWSSLG----- 339

QY 297 DFIARDLGPTLANSTHNVRLMLDDQRLLLPHWAKVLTDPKAAKYVHCVHVVYL--- 353
DB 340 --VSEGSQP-----AETANPVNASNIM---AFHFTAAASH 370

QY 354 --DFLAPAKATLGTETRLPNTMLFASACVGSKFWEQSVRLGSDRGWQYSHSIITNLL 411
DB 371 RDNLL---NALREASELP---VFVTE---FGTETY-----TGDGANDFQMDRYIDLMA 416

QY 412 YHVVGTWTDNLA-----LNP-----EGGP-----NWRNRFVDS 439
DB 417 ERKIGWTKWNYSDFRSGAVFGFGTCASGGPWSGSSLKASGQWVRSKLOS 466

RESULT 7
GALC_MOUSE
ID GALC_MOUSE STANDARD; PRT; 668 AA.
AC P54818; O35151;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
```

15-JUL-1999 (Rel. 38, Last annotation update)
 Galactocerebrosidase precursor (EC 3.2.1.46) (GALCERASE)
 (Galactosylceramide) (Galactosylceramide beta-galactosidase)
 (Galactocerebroside beta-galactosidase).
 GN GALC.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=96365697; PubMed=8769874;
 RA Sakai N., Inui K., Tatsumi N., Fukushima H., Nishigaki T.,
 RA Tanike M., Nishimoto J., Tsukamoto H., Yanagihara I., Ozono K.,
 RA Okada S.;
 RT "Molecular cloning and expression of cDNA for murine
 RT galactocerebrosidase and mutation analysis of the twitcher mouse, a
 RT model of Krabbe's disease.";
 RT J. Neurochem. 66:1118-1124 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Luzi P., Victoria T., Wenger D.A.;
 RT "Genomic organization of the mouse galactocerebrosidase (GALC) gene.";
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HYDROLYSES THE GALACTOSE ESTER BONDS OF
 CC GALACTOSYLKERAMIDE, GALACTOSYLSPHINGOSINE, LACTOSYLKERAMIDE, AND
 CC MONOGALACTOSYLGLYCERIDE. ENZYME WITH VERY LOW ACTIVITY
 CC RESPONSIBLE FOR THE LYSOSOMAL CATABOLISM OF GALACTOSYLKERAMIDE, A
 CC MAJOR LIPID IN MYELIN, KIDNEY AND EPITHELIAL CELLS OF SMALL
 CC INTESTINE AND COLON (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: D-galactosyl-N-acylsphingosine + H(2)O = D-
 CC galactose + N-acylsphingosine.
 CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).
 CC -1- DISEASE: DEFECTS IN GALC ARE THE CAUSE OF THE 'TWITCHER'
 CC PHENOTYPE, AN AUTOSOMAL RECESSIVE LEUKODYSTROPHY SIMILAR TO THE
 CC HUMAN DISEASE (KRABBE DISEASE). THIS DEFICIENCY RESULTS IN THE
 CC INSUFFICIENT CATABOLISM OF SEVERAL GALACTOLIPIDS THAT ARE
 CC IMPORTANT IN THE PRODUCTION OF NORMAL MYELIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 59 OF GLYCOSYL HYDROLASES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D38537; BAA07560.1; -
 CC EMBL; AF003886; AAB71823.1; JOINED.
 CC EMBL; AF003870; AAB71823.1; JOINED.
 CC EMBL; AF003871; AAB71823.1; JOINED.
 CC EMBL; AF003872; AAB71823.1; JOINED.
 CC EMBL; AF003873; AAB71823.1; JOINED.
 CC EMBL; AF003874; AAB71823.1; JOINED.
 CC EMBL; AF003875; AAB71823.1; JOINED.
 CC EMBL; AF003876; AAB71823.1; JOINED.
 CC EMBL; AF003877; AAB71823.1; JOINED.
 CC EMBL; AF003878; AAB71823.1; JOINED.
 CC EMBL; AF003879; AAB71823.1; JOINED.
 CC EMBL; AF003880; AAB71823.1; JOINED.
 CC EMBL; AF003881; AAB71823.1; JOINED.
 CC EMBL; AF003882; AAB71823.1; JOINED.
 CC EMBL; AF003883; AAB71823.1; JOINED.
 CC EMBL; AF003884; AAB71823.1; JOINED.
 CC EMBL; AF003885; AAB71823.1; JOINED.
 CC MGD; MGI:95636; Galt.
 CC InterPro: IPR001286; GH_59.
 CC Pfam: PF02057; Glyco_hydro_59; 1.
 CC PRINTS; PR00850; GLHYDRLASE59.
 CC Hydrolase; Glycosidase; Glycoprotein; Signal.
 CC SIGNAL 1 26 BY SIMILARITY.

FT CHAIN 27 668 GALACTOCEREBROSIDASE.
 FT CARBOHYD 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 360 G -> A (IN REF. 2).
 SQ SEQUENCE 668 AA; 75502 MW; 3BAD82D9E7C2842 CRC64;
 Query Match 3.5%; Score 99; DB 1; Length 668;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 81; Conservative 51; Mismatches 133; Indels 126; Gaps 23;
 QY 119 GFGGAMTDAALNTIALSP-PAONLLKSYFSEE-GIGYNIIRVPM-SCDFSIRTY 172
 DB 41 GIGAVGGGATSLRLVNYPEYSEILDYLFKPNFGASLHLAVEIGGQTTDGTPEPSH 100
 QY 173 TYADTPDD-FQLHNFSLPEEDTKIKPLIHRALQALQAPVSLASPTSTWLKTNAGVN 231
 DB 101 MHVELDENYFRGYEWLMKEAKRNPDI--LMGLPWSFPGL-- 141
 QY 232 GKSLKQPGDIYHQTW-----ARYEVKF-LDAYAEHKLQFVAVTAENEFSAAGLSG 282
 DB 142 KGKF-----SWPVNQLQTATYVYVWILGAKHYHDLIDYI-----GIWNE 182
 QY 283 YPQCLGFTPEHQRDFTARDIGPTLANSTHNVRLMLDQRLLLPHWAKV---VLTDP 339
 DB 183 RPFDA-----NYI-KELRMLDYQGLQVRVLIASDNL-----WEPISLSLLDQE 226
 QY 340 AAKYVHGIAVHWYLDLAPAKATLGETHRFPNTMLFASACVSGKFW-----EQSV 391
 DB 227 LMKVVDVIGAH-----YPGTYTVNNAKMSCKLWSSDFSTINSY 267
 QY 392 RLGSWDR--GMQYSHSIITNLLYHVGVWTDNIALN-----PEGGP- 436
 DB 268 GAGCWSRILNONGINGMTSTI-----AMNVASYEELPYGRSLMTAQBPWSGHYV 320
 QY 437 VDSPLIIVDITKDTFYKQPMFYHL---GHFSK 464
 DB 321 VASPIWVS-AHTTQTPGWYLYKTVGHLEK 350

RESULT 8
 GALC_MACMU STANDARD; PRT; 669 AA.
 AC 002791;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Galactocerebrosidase precursor (EC 3.2.1.46) (GALCERASE)
 DE (Galactosylceramide) (Galactosylceramide beta-galactosidase)
 DE (Galactocerebroside beta-galactosidase).
 GN GALC.
 OS Macaca mulatta (Rhesus macaque).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
 CC Cercopitheciinae; Macaca.
 CC NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336058; PubMed=9192853;
 RA Luzi P., Rafi M.A., Victoria T., Baskin G.B., Wenger D.A.;
 RT "Characterization of the rhesus monkey galactocerebrosidase (GALC)
 RT cDNA and gene and identification of the mutation causing globoid cell
 RT leukodystrophy (Krabbe disease) in this primate.";
 RL Genomics 42:319-324 (1997).
 CC -1- FUNCTION: HYDROLYSES THE GALACTOSE ESTER BONDS OF
 CC GALACTOSYLKERAMIDE, GALACTOSYLSPHINGOSINE, LACTOSYLKERAMIDE, AND
 CC MONOGALACTOSYLGLYCERIDE. ENZYME WITH VERY LOW ACTIVITY
 CC RESPONSIBLE FOR THE LYSOSOMAL CATABOLISM OF GALACTOSYLKERAMIDE, A
 CC MAJOR LIPID IN MYELIN, KIDNEY AND EPITHELIAL CELLS OF SMALL
 CC INTESTINE AND COLON (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: D-galactosyl-N-acylsphingosine + H(2)O = D-
 CC galactose + N-acylsphingosine.
 CC -!- SUBCELLULAR LOCATION: Lysosomal (By similarity).
 CC -!- DISEASE: DEFECTS IN GALC ARE THE CAUSE OF GLOBOID CELL
 CC LEUKODYSTROPHY (GLD) (OR KRAHE DISEASE). THIS DEFICIENCY RESULTS
 CC IN THE INSUFFICIENT CATABOLISM OF SEVERAL GALACTOLIPIDS THAT ARE
 CC IMPORTANT IN THE PRODUCTION OF NORMAL MYELIN.
 CC -!- SIMILARITY: BELONGS TO FAMILY 59 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

EMBL; U87477; AAB58575.1; JOINED.
 EMBL; U87462; AAB58575.1; JOINED.
 EMBL; U87463; AAB58575.1; JOINED.
 EMBL; U87464; AAB58575.1; JOINED.
 EMBL; U87465; AAB58575.1; JOINED.
 EMBL; U87466; AAB58575.1; JOINED.
 EMBL; U87467; AAB58575.1; JOINED.
 EMBL; U87468; AAB58575.1; JOINED.
 EMBL; U87469; AAB58575.1; JOINED.
 EMBL; U87470; AAB58575.1; JOINED.
 EMBL; U87471; AAB58575.1; JOINED.
 EMBL; U87472; AAB58575.1; JOINED.
 EMBL; U87473; AAB58575.1; JOINED.
 EMBL; U87474; AAB58575.1; JOINED.
 EMBL; U87475; AAB58575.1; JOINED.
 EMBL; U87476; AAB58575.1; JOINED.
 EMBL; U87628; AAB58576.1; JOINED.
 InterPro: IP001286; GH_59.
 Pfam: PF02057; Glyco_hydro_59; 1.
 PRINTS; PR00850; GLHYDRIASE59.
 Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 668
 FT CARBOHYD 127 127
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 363 363
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 387 387
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 435 435
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 543 543
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 586 586
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 669 AA; 75206 MW; F09169EDC2B66C07 CRC64;

Query Match 3.4%; Score 97.5; DB 1; Length 669;
 Best Local Similarity 20.3%; Pred. No. 2.7; Mismatches 153; Indels 135; Gaps 24;
 Matches 87; Conservative 53;
 QY 119 GFGAMTDAAALNIALSP-PAQNLLKVSFSEB-GIGYNIIRVPM-----SCDFSIRTY 172
 DB 41 GIGAVSGGATSRLLVNYVPEYRSDLDYLFKPNFGASHLKVEIGGQGTDTGTEPSH 100
 QY 173 T-YADTPDDFQHLNFSLPEDTKLPIHLRALQALQRP-VSLASPTWTSPTWLKNGAV 230
 DB 101 MHYALDENTFRGYEWMKAEAK-----RNPNTILIGLPWSFPGWL----- 141
 QY 231 NGKG-----SLKQPGDIYHOTWARYFVKFLDAYAEHLQFVAVTAENPESAGLLSGYPFQ 286
 DB 142 -GKGFDPVYVNLQLTAYVVTW-----IVGAKRYHDLDDIDYIGIWNERS----- 184
 QY 287 CLGFTPEHQRDFIARDIGPTLANSTHNRVLLMDQRLLEPHWAKV---VLTOPEAKY 343
 DB 185 -----YNANYI-KILRKLNSOGLQVKKIIASDNL-----WESISAALLDAELFKV 230
 QY 344 VHGIAVHWLDFLAPAKATLGETHRLPNTMLFASACVGSKEF-----EOSVRLGS 395
 DB 231 VDVIAGH-----YPTGHSVKDARLTGKKLWSSSEDFSLNSDTGAGC 271

QY 396 WDR--GMOYSHIITNLLYHVGVGTWNALN-----PEG-----GPNVVRNF-VDSP 440
 DB 272 WGRILNQVNGVYMTSTI-----AWNIVASYEQLPYGRGLMTAQEPWGSYHYVVEP 324
 QY 441 IIVDITKDTFYKQPMFYHL---GHFSKFFIPEGSQVRLVVASQKNDLDAVALMHPDGSADV 497
 DB 325 VVWS-AHTTPTOPGWYLTGVHLEK---GGSY-----VALTDGLGNLTI 366
 QY 498 VVLRSSK 505
 DB 367 LIETMSHK 374

RESULT 9

IMD_ARTGO STANDARD; PRT; 641 AA.
 ID Q44052;
 AC Q44052;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Isomalto dextranase precursor (EC 3.2.1.94) (Glucan 1,6-alpha-
 DE isomaltosidase) (Exo-isomaltohydrolase).
 GN IMD.
 OS Arthrobacter globiformis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Micrococccineae; Micrococccaceae; Arthrobacter.
 OX NCBI_TaxID=1665;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=T6;
 RX MEDLINE=95095945; PubMed=8002600;
 RA Iwai A., Ito H., Mizuno T., Mori H., Matsui H., Honma M., Okada G.,
 RA Chiba S.;
 RT "Molecular cloning and expression of an isomalto-dextranase gene from
 RT Arthrobacter globiformis T6."
 RL J. Bacteriol. 176:7730-7734(1994).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
 CC in polysaccharides so as to remove successive isomaltose units
 CC from the non-reducing ends of the chains.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

EMBL; D30761; BAA06424.1;
 DR InterPro: IPR000111; Glyco_hydro_GHD.
 DR PRODOM: PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; FALSE_NEG.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 39
 FT CHAIN 40 641
 SQ SEQUENCE 641 AA; 69764 MW; B7F8F2F8F4D88350 CRC64;

Query Match 3.4%; Score 97; DB 1; Length 641;
 Best Local Similarity 18.9%; Pred. No. 2.8; Indels 212; Gaps 28;
 Matches 99; Conservative 55;
 QY 105 LLLTPEQFKQVKFGGAMTDAAALNI-----LALSPQAQNLKLSYFSEBGI---YN 156
 DB 9 LITGSAATLAYALMGSAQAATAVTARPCVPTAAPPRLASRNSVTRSGAGPRYWN 68
 QY 157 IIRVPMASCDFSIRTYTADTPDDPQLHNSLPEDTKLPIHLRALQALQRPVSLAS 216
 DB 69 I-----YGSFPP-----HNAPIPENEWKANIDWL--AGNFADFGYDIAC- 105
 QY 217 PWTSPWLKNGAVNGKSLKQPGDIYHOTWARYFVKFLDAYAEHLQ-----FWAV 269

Db 106 ---TDGWIESSRTTNGVITSY-NDSSQHDWA-YWANYL---AARKMKLVYINPLWVH 157
QY 270 TAENPSAGLLSGYFPQCLG--FTPEHORDFTARDLGTLANSHHNVRLMLDDQRLLL 327
Db 158 RAAVEDASKTVLGRPDVADIADLVVP---GDFEFARDIGN-----OLYLD----- 199
QY 328 PHWAKVVLDPAAKYVHUGIAVHW-----YL--DFLA-----PAKATIGETH-----RLF 370
Db 200 -----VTKSAKEYVQGVYFVFDLGVYLRIDFLSWYEDGRDANIGVNAHPGRANY 252
QY 371 PNTMLFASAC-----VGSKEWESQVRL-----GSDR--GMQYSHS 405
Db 253 ELALSINEAAGEDMEVSLVPMFQDGSALANGDLVRINADKGGWRLSGMRQN-- 310
QY 406 IITNLLYHVVGWTD--WNLALNPEGGPNVVRNF----- 436
Db 311 -----WODAW-----PNWANPCGFTGWSHRNGRQLILDGDFMRASTFA 350
QY 437 -----VDSPI-----IVDITKDTFYKQPMFYHLGHF 462
Db 351 SDEERTMNLVMAAGSPLAIADTYQIGNNAWVYTNKEVLQNLNADGLVGKPLYSATPF 410
QY 463 SKFIPESQVGLVASQKNDLDAVLMHPDGSVVVVLNRSKD 506
Db 411 SK--DPGR-----DTERWAGQLPDGSGWVAFNRSDTE 442

RESULT 10
PTPO_RAT STANDARD; PRT: 1711 AA.
AC Q64612;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Osteosticular protein tyrosine phosphatase precursor (EC 3.1.3.48) (OST-PTP).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Osteosarcoma;
RX MEDLINE=95074080; PubMed=7527035;
RA Mauro L.J., Olmsted E.A., Skrobacz B.M., Mourey R.J., Davis A.R., Dixon J.E.;
RT Identification of a hormonally regulated protein tyrosine phosphatase associated with bone and testicular differentiation.";
RL J. Biol. Chem. 269:30659-30667(1994).
CC -!- FUNCTION: MAY FUNCTION IN SIGNALING PATHWAYS DURING BONE REMODELING, AS WELL AS SERVE A BROADER ROLE IN CELL INTERACTIONS ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS. OPTIMAL PH FOR PHOSPHATASE ACTIVITY IS 5.6.
CC ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing. A presumed alternate transcript of 4.8-5.0 kilobases, which may lack PTP domains, is present in proliferating osteoblasts, but not detectable at other stages.
CC -!- TISSUE SPECIFICITY: BONE AND TESTIS. IN THE LATTER, RESTRICTED TO THE BASAL PORTION OF THE SEMINIFEROUS TUBULE.
CC -!- DEVELOPMENTAL STAGE: UP-REGULATED IN DIFFERENTIATING CULTURES OF PRIMARY OSTEOBLASTS AND DOWN-REGULATED IN LATE STAGE MINERALIZING CULTURES. IN TESTIS, EXPRESSION IS HIGHEST BETWEEN STAGES I AND VII WHEN MATURING SPERMATIDS REMAIN BURIED WITHIN THE SERTOLI EPITHELIUM.
CC -!- INDUCTION: BY PARATHYROID HORMONE AND CYCLIC AMP ANALOGS.
CC -!- PTM: THE CYTOPLASMIC DOMAIN CONTAINS POTENTIAL PHOSPHORYLATION SITES.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -!- SIMILARITY: CONTAINS 10 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL: L36884; AAA63911.1; -
HSSP: P18052; LYFO.
DR InterPro: IPR000340; DS_phosphatase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FN_III_repeat.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_PP.
DR Pfam: PF00041; fn3; 7
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00014; ENTPEI11.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPc; 1.
DR SMART: SM00012; PTPc_DSPc; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00036; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolyase; Transmembrane; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1711 OSTEOESTICULAR PROTEIN TYROSINE PHOSPHATASE.
FT FT 18 1074 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1075 1095 POTENTIAL.
FT DOMAIN 1096 1711 CYTOPLASMIC (POTENTIAL).
FT FT 124 FIBRONECTIN TYPE-III 1.
FT FT 125 215 FIBRONECTIN TYPE-III 2.
FT FT 216 303 FIBRONECTIN TYPE-III 3.
FT FT 304 392 FIBRONECTIN TYPE-III 4.
FT FT 393 470 FIBRONECTIN TYPE-III 5.
FT FT 471 562 FIBRONECTIN TYPE-III 6.
FT FT 563 652 FIBRONECTIN TYPE-III 7.
FT FT 653 741 FIBRONECTIN TYPE-III 8.
FT FT 742 830 FIBRONECTIN TYPE-III 9.
FT FT 831 921 FIBRONECTIN TYPE-III 10.
FT FT 1150 1418 PROTEIN-TYROSINE PHOSPHATASE 1.
FT FT 1419 1711 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 1350 1350 BY SIMILARITY.
FT FT 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 663 663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT 982 982 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1711 AA; 187292 MW; 0F04D2D1A47A18A0 CRC64;

Query Match 3.4%; Score 96; DB 1; Length 1711;
Best Local Similarity 19.6%; Pred. No. 13;
Matches 107; Conservative 59; Mismatches 159; Indels 220; Gaps 31;

QY 131 NTLALSPAQNLLKSYSEIGYNIIRVPMASCDFSIRTYTYADT--PDFOLHNFSL 188
Db 519 DIVDLGPDTLSTLKS-----VP-GSC-YTVSAWAGNLDSDSQIHSCTR 564

QY 189 PREDTKLKIPLHRLALQAPVSLASPTWTSPTWLKTNAGVNGK 234
 DB 565 PAPPNTL-----SLGFAHOPAALKAS-WYHPP-----GGRDAFHLLRLRLPL 608
 QY 235 -SLGPGDGIYHOPWARYFVKFDAYAEHLQ---FWAVTAENEPSAGLLSGPFOCLGF 290
 DB 609 ESEKVLPREAQNFWAO-----LTACEPOVOLSTLWG--SERSSA-----NATGW 653
 QY 291 TPEHOREFIARDLGPPLANTSHINVELLMDDORLLPLHAKV-----VLTDPAAK 342
 DB 654 TTPS-----APTLNVNTS-----DAPTOLOVSWAHVPGGRSRYQVTLYOESTR 696
 QY 343 VVHGI-----AVHWYLDLAPAKATGETHRLPNTMLFASB- 379
 DB 697 TATSMGPKEDGTSFLGLTPGTKYKVEISWAGPLTYAAANVSANTYPLIPNELLVSMQA 756
 QY 380 -----AC-----VSGKFEQSVRLGS---WDRGMQYSHSI----- 406
 DB 757 GSAVVNLAWPSGLPGOGACHAQLSDAGHLSWEOPLKQLQELFMLRLDTPGHTISMSVRCR 816
 QY 407 -----ITNLLYH-----VVGWT-----DWNLA-----LNPEGSPN 431
 DB 817 AGPLQASTHLWLSVEPGPVEDVLCHEATVYALNLTMTMPAGDVDCVLYVVERLVPGGTH 876
 QY 432 W-----VRNFVDSPIVDITDTFKOPMFVHLGHFSKETPEGSORVGLVASOKNDLDAVAL 488
 DB 877 FVEQVNTSGDALLPLNLTPTSYRLSLIV-LGRNSRW-----SRAVSLVCS-----TSAEA 926
 QY 489 MHP-----DGSAAVVVNLRSKDVPLIKDPV-----GFLETIS-----PGYSI 528
 DB 927 WHPELAEPQVVELGCMGTVMRCMFGD-----DQIQWYGIITATINMTLQAQSPREA 980
 QY 529 HTYLM 533
 DB 981 INYTW 985
 RESULT 11
 HMZ2-ARATH STANDARD; PRT; 512 AA.
 AC 004921; 023623;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ferriochelatease II, chloroplast precursor (EC 4.99.1.1) (Protoheme
 ferri-lyase) (Heme synthetase).
 GN AT2G30390 OR T06B20-24 OR T09D09.20
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Seedling;
 RX MEDLINE=98426386; PubMed=9753778;
 RA Chow K.-S., Singh D.P., Walker A., Smith A.G.;
 RT "Two different genes encode ferriochelatease in Arabidopsis: mapping,
 RL expression and subcellular targeting of the precursor proteins.";
 RN Plant J. 15:531-541(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.I.,
 RA Moffat K.S., Cronin L.A., Shen W., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman H.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana.";
 RL Nature 402:761-768(1999).
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=97354168; PubMed=9210462;
 RA Roper J.M., Smith A.G.;
 RT "Molecular localisation of ferriochelatease in higher plant
 chloroplasts.";
 RL Eur. J. Biochem. 246:32-37(1997).
 CC -!- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX.
 CC MAY HAVE A ROLE IN DEALING WITH OXIDATIVE STRESS.
 CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) -> protoheme + 2 H(+).
 CC -!- PATHWAY: Protoheme biosynthesis; last step.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST (THYLAKOID AND ENVELOPE
 MEMBRANES).
 CC -!- SIMILARITY: BELONGS TO THE FERROCHELATEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to a license@sib-sib.ch).
 CC
 DR EMBL; U93215; AAB63095.1; -;
 DR EMBL; Y13156; CAAY3614.1; -;
 DR HSSP; P32396; IFJ1.
 DR InterPro; IPR001015; Ferriochelatease.
 DR Pfam; PF00762; Ferriochelatease; 1.
 DR ProDom; PD002792; Ferriochelatease; 1.
 DR TIGRFAMs; TIGR00109; hemH; 1.
 DR PROSITE; PS00534; FERROCHELATEASE; 1.
 KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron; Chloroplast;
 KW Transit peptide.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT CHAIN ? 512 FERROCHELATEASE II.
 FT DOMAIN 9 19 POLY-SER.
 FT DOMAIN 23 26 POLY-PRO.
 FT DOMAIN 80 83 POLY-SER.
 FT DOMAIN 90 93 POLY-SER.
 FT CONFLICT 156 159 SKEG -> ARR (IN REF. 1).
 FT SEQUENCE 512 AA; 56618 MW; 521BIFCS96EB4A7 CRC64;
 SQ
 Query Match 3.3%; Score 95; DB 1; Length 512;
 Best Local Similarity 18.9%; Pred. No. 3;
 Matches 102; Conservative 72; Mismatches 172; Indels 194; Gaps 26;
 QY 67 PPTPALGT-----FSRYESTRSGRRMELSMGPIQANHTGTGLLTLPBQ 112
 DB 25 PLLLPQLSNDQSRSVYVHCTRLPFEAFATSSNLLGKHSPLRA-----ALVTSNP-- 76
 QY 113 KFOVKFGGAMTDAALNIALSPQAQNLKSYFSEEG-IGNIIRV--PMASDFSI 169
 DB 77 -----LNISSSSVISDAISSSVITDDAKIGVLLNLGGP----- 111
 QY 170 RTYTYADTPDFOLHNFSLPEEDTKLKIPLHRLALQAPVSLASPTWTSPTWLKTNCA 229
 DB 112 -----ETLDDVQPFELNLFADPDIIRLPPV---FQLQKPLAQFISVARAKSKEGYAS 162
 QY 230 VNGKSLKGGOPGDIVHCTWARYFVFLDAYAEH-KLOFWAVTAENEPSAGLLSGPFOCL 288
 DB 163 IGG-----GSP--LRHIT-----DQAEELKCKWEKNPKVAKVYVGMRYWHPF--- 203
 QY 289 GFTPEHORDFIARD-----LGPTIANSTH-HNVRLL-----MLDQRLLLPH 329
 DB 204 ---TEEAIEQIKRGDTIKLVLPPLYPQFSISTSGSLRLLERIFREDEYLVNMQHTVIPS 260
 QY 330 W-----AKVVLTDPAAKY-----VHGIAVHWYLDLFLAPAKATLGE----- 365
 DB 261 WYQREGYIKAMANLIQSELGKFSNPQVVIFFSAHGVPFLAYVEEAGDPYKAEWECVLDI 320
 QY 366 ---THRLFPNTMLFASACVGSKEWQSVRLGSWDRGMQYSHSIITNLLYHVVGWTDN 421

```

Db 321 MEELDKRKITNAYTAYQSRVGPVEMLK-----PYTERAITEI----- 358
QY 422 LALNPEGGNWVRNFVDSPI-----IVDTKDTYK-----Q 453
Db 359 -----GKKGVENLLAVIPSPVSEHETLEI--DVEYKELAKSGIKKNGRVPALGTE 409
QY 454 PMFYHLGHSKIPESQVGLVASKNDLDAVALMHPDGS--VVVVLNRSSKDVPLTI 511
Db 410 PMF--ISLADAVVSLPYVGAMA--VSNLEARQSLVPLGSEELLATYSQRRELPAVP 465

RESULT 12
C11A_DASAM
ID C11A_DASAM STANDARD; PRT; 514 AA.
AC Q92045;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Cytoschrome P450 11A1, mitochondrial precursor (EC 1.14.15.6) (CYPX1A1)
DE (P450(SCC)) (Cholesterol side-chain cleavage enzyme) (Cholesterol
DE desmolase) (Fragment).
GN CYP11A1.
OS Dasyatis americana (Southern stingray).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hymnosqualea; Pristiogalea; Batoidae;
OC Myliobatiformes; Myliobatidae; Dasyatidae; Dasyatis.
OX NCBI_TaxID=50411;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestinal gland;
RX MEDLINE=97225805; PubMed=9073075;
RA Nunez S., Trant J.M.;
RT "Isolation of the putative cDNA encoding cholesterol side chain
RT cleavage cytochrome P450 (CYP11A) of the southern stingray (Dasyatis
RT americana).";
RL Gene 187:123-129(1997).
CC -1- FUNCTION: CATALYZES THE SIDE-CHAIN CLEAVAGE REACTION OF
CC CHOLESTEROL TO PREGNENOLONE.
CC -1- CATALYTIC ACTIVITY: Cholesterol + reduced adrenal ferredoxin +
CC O(2) -> pregnenolone + 4-methylpentanal + oxidized adrenal
CC ferredoxin + H(2)O.
CC -1- PATHWAY: INITIAL RATE-LIMITING REACTION IN THE SYNTHESIS OF
CC VARIOUS STEROID HORMONES.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U63299; AAC60095.1; -
CC DR HSSP; P00189; 1SCC.
CC DR InterPro: IPR001128; Cytochrome_P450.
CC DR Pfam: PF00067; P450.1.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Membrane; Heme; Mitochondrion;
CC KW Steroidogenesis; Transit peptide.
CC FT NON_TER 1 1
CC FT TRANSIT <1 39 MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN 40 514 CYTOCHROME P450 11A1.
CC FT BINDING 461 461 HEME (BY SIMILARITY).
CC FT SEQUENCE 514 AA; 59863 MW; D0F03E25D5534FB2 CRC64;
SQ
Query Match 3.3%; Score 95; DB 1; Length 514;
Best Local Similarity 20.1%; Pred. NO. 3;
Matches 92; Conservative 51; Mismatches 136; Indels 178; Gaps 24;
QY 168 SIRTYYADTPDDFQLHNSLPEDTKL-----KIPLIHRALQLAQRVPVSLIA 215

```

```

Db 5 SLSASTVA-----QRGSFTTPHEDETLFPHRNHSVTSERIPSEQTLKSLDIP----- 53
QY 216 SPWTSPTWL-----KTNGAVNGKSGLKQPGDIYHTWARYEVFKFLDAYAEHKLOFWAV 269
Db 54 GNMWK-NLNLVNYFWRSNLNNA-----HQMLDNENKYGPIYREKIAYESI 100
QY 270 TAENEPAGLL--SGYPFOCLGFTPEH-----QRDFIARDLGPTLANSTHNNVRLLM 319
Db 101 NIINPADAVIMNKSEGF-----PKRIEMAPVAVYRDLRKENYGVQLLNGENWKRTLI 154
QY 320 LDDQ-----RLLLPHWAKVVL-----TDPEAAKYVHGIAPHWY- 352
Db 155 LNSIFAQSSIORLVLPLNEVLDVFSMVHKEVESKSDYWKTDLTNDLFKLALEWICYI 214
QY 353 -----LDLF-----APAKATLGETHRLFPNT--MLFASEA---CVGSKFWEQSVRLGWS 396
Db 215 LYGERLLOLQRYNKAPQK-FIDSIATMFHSTPMLYVPSLSKINSKIQQHV--GSW 271
QY 397 D-----RGMQYSHS---IITNLYH----- 413
Db 272 DNIPEHADTYLKAYROFQOGSKNEHAFPGVLTLLQGLALPEDIIRASIIDVMSGAIDT 331
QY 414 ---VVGTDNWLALNP-----EGGNWYRNFDSPILIVDITKDTFYKQ 453
Db 332 TSTTVHWMYELAKHPHIQKNRSEIMEAHOKTEGDP--VKMLKSVPLLCVKVKTURLY 389
QY 454 PMFYHLGHF-----SKFIPEGS-QRVGLVASQKN 481
Db 390 PVAISIORYLNEDIVLQNYHIPAGTLVQLGLYAMGRN 426

RESULT 13
GALC_HUMAN
ID GALC_HUMAN STANDARD; PRT; 669 AA.
AC P54803;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactocerebroside precursor (EC 3.2.1.46) (GALCERASE)
DE (Galactosylceramidase) (Galactosylceramide beta-galactosidase)
DE (Galactocerebroside beta-galactosidase).
GN GALC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta, and Skin fibroblast;
RX MEDLINE=94128088; PubMed=8297359;
RA Sakai N., Inui K., Fujii N., Fukushima H., Nishimoto J.,
RA Yanagihara I., Isegawa Y., Iwamatsu A., Okada S.;
RT "Krabbe disease: isolation and characterization of a full-length cDNA
RT for human galactocerebroside.";
RL Biochem. Biophys. Res. Commun. 198:485-491(1994).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-59 AND 436-454.
RC TISSUE=Brain, and Testis;
RX MEDLINE=94108435; PubMed=8281145;
RA Chen Y.O., Rafi M.A., de Gala G., Wenger D.A.;
RT "Cloning and expression of cDNA encoding human galactocerebroside,
RT the enzyme deficient in globoid cell leukodystrophy.";
RN Hum. Mol. Genet. 2:1841-1845(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324938; PubMed=7601472;
RA Luzi P., Rafi M.A., Wenger D.A.;
RT "Structure and organization of the human galactocerebroside (GALC)
RT gene.";
RL Genomics 26:407-409(1995).
RN [4]
RP SEQUENCE FROM N.A.

```

MEDLINE=98094242; PubMed=9434153;
 RA Takai N., Fukushima H., Inui K., Fu L., Nishigaki T., Yanagihara I.,
 RA Tatsumi N., Ozono K., Okada S.;
 RT "human galactocerebroside gene: promoter analysis of the 5'-flanking
 RT region and structural organization."; *Hum. Genet.* 100:11-17(1997).
 RL Biochim. Biophys. Acta 1395:62-67(1998).
 RN [5]
 RN SEQUENCE OF 27-45 AND 436-454, AND CHARACTERIZATION.
 RC TISSUE=Urine;
 RA MEDLINE=94002192; PubMed=8399327;
 RA Chen Y.-O., Wenger D.A.;
 RT "Galactocerebroside from human urine: purification and partial
 RT characterization."; *Biochem. Biophys. Res. Commun.* 191:111-115(1993).
 RL Biochim. Biophys. Acta 1170:53-61(1993).
 RN [6]
 RN REVIEW ON GLD MUTATIONS.
 RC MEDLINE=97478285; PubMed=9338580;
 RA Wenger D.A., Rafi M.A., Luzzi P.;
 RT "Molecular genetics of Krabbe disease (globoid cell leukodystrophy):
 RT diagnostic and clinical implications."; *Hum. Mutat.* 10:268-279(1997).
 RN [7]
 RN VARIANTS GLD ALA-302 AND GLY-550.
 RC MEDLINE=96121583; PubMed=8595408;
 RA Tatsumi N., Inui K., Sakai N., Fukushima H., Nishimoto J.,
 RA Yanagihara I., Nishigaki T., Tsukamoto H., Fu L., Tanike M.,
 RA Okada S.;
 RT "Molecular defects in Krabbe disease."; *Hum. Mol. Genet.* 4:1865-1868(1995).
 RL [8]
 RN VARIANTS GLD ASN-528 AND SER-583.
 RC MEDLINE=96198195; PubMed=8786069;
 RA Rafi M.A., Luzzi P., Zlotogora J., Wenger D.A.;
 RT "Two different mutations are responsible for Krabbe disease in the
 RT Druse and Moslem Arab populations in Israel."; *Hum. Genet.* 97:304-308(1996).
 RL [9]
 CC FUNCTION: HYDROLYSES THE GALACTOSE ESTER BONDS OF
 CC GALACTOSYLKERAMIDE, GALACTOSYLSPHINGOSINE, LACTOSYLKERAMIDE, AND
 CC MONOGALACTOSYLGLYCERIDE. ENZYME WITH VERY LOW ACTIVITY
 CC RESPONSIBLE FOR THE LYSOSOMAL CATABOLISM OF GALACTOSYLKERAMIDE, A
 CC MAJOR LIPID IN MYELIN, KIDNEY AND EPITHELIAL CELLS OF SMALL
 CC INTESTINE AND COLON. HAS AN OPTIMAL PH BETWEEN 4.0 AND 4.4.
 CC ACTIVITY IS LOST WHEN HEATED AT 52 DEGREES CELSIUS FOR FIVE
 CC MINUTES.
 CC [10]
 CC CATALYTIC ACTIVITY: D-galactosyl-N-acylsphingosine + H(2)O = D-
 CC galactose + N-acylsphingosine.
 CC [11]
 CC SUBCELLULAR LOCATION: Lysosomal.
 CC [12]
 CC ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF THE LONG FORM.
 CC [13]
 CC TISSUE SPECIFICITY: HIGHEST LEVEL OF ACTIVITY IN TESTES COMPARED
 CC TO BRAIN, KIDNEY, PLACENTA AND LIVER. CAN ALSO BE FOUND IN URINE.
 CC [14]
 CC DISEASE: DEFECTS IN GALC ARE THE CAUSE OF GLOBOID CELL
 CC LEUKODYSTROPHY (GLD) (OR KRABBE DISEASE). THIS AUTOSOMAL RECESSIVE
 CC DISORDER DEFICIENCY RESULTS IN THE INSUFFICIENT CATABOLISM OF
 CC SEVERAL GALACTOLIPIDS THAT ARE IMPORTANT IN THE PRODUCTION OF
 CC NORMAL MYELIN. CLINICALLY THE MOST FREQUENT FORM IS THE INFANTILE
 CC FORM. MOST PATIENTS (90%) PRESENT BEFORE SIX MONTHS OF AGE WITH
 CC IRRITABILITY, SPASTICITY, ARREST OF MOTOR AND MENTAL DEVELOPMENT,
 CC AND BOUTS OF TEMPERATURE ELEVATION WITHOUT INFECTION. THIS IS
 CC FOLLOWED BY MYOCLONIC JERKS OF ARMS AND LEGS, OPHTHOMOSUS,
 CC HYPERTONIC FITS, AND MENTAL REGRESSION WHICH PROGRESSES TO A
 CC SEVERE DEGENERATE CONDITION WITH NO VOLUNTARY MOVEMENTS AND DEATH
 CC FROM RESPIRATORY INFECTIONS OR CEREBRAL HYPERPEREIA BEFORE 2
 CC YEARS OF AGE. HOWEVER, A SIGNIFICANT NUMBER OF CASES WITH LATER
 CC ONSET, PRESENTING WITH UNEXPLAINED BLINDNESS, WEAKNESS AND/OR
 CC PROGRESSIVE MOTOR, AND SENSORY NEUROPATHY THAT CAN PROGRESS TO
 CC SEVERE MENTAL INCAPACITY AND DEATH, HAVE BEEN IDENTIFIED.
 CC [15]
 CC SIMILARITY: BELONGS TO FAMILY 59 OF GLYCOSYL HYDROLASES.
 CC [16]

```

FT VARIANT 247 247 A -> T (IN GLD).
FT VARIANT 268 268 /FtId=VAR_003388.
FT VARIANT 270 268 G -> S (IN GLD).
FT VARIANT 270 270 /FtId=VAR_003389.
FT VARIANT 279 270 G -> D (IN GLD).
FT VARIANT 279 279 /FtId=VAR_003390.
FT VARIANT 287 287 N -> T (IN GLD).
FT VARIANT 287 287 /FtId=VAR_003391.
FT VARIANT 302 302 S -> F (IN GLD).
FT VARIANT 302 302 /FtId=VAR_003392.
FT VARIANT 380 380 P -> A (IN GLD).
FT VARIANT 380 380 /FtId=VAR_003393.
FT VARIANT 380 380 R -> W (IN GLD, BILATERAL CHERRY RED
FT VARIANT 380 380 SPOTS).
FT VARIANT 380 380 /FtId=VAR_003394.

Query Match 3.3%; Score 94.5; DB 1; Length 669;
Best Local Similarity 20.3%; Pred. No. 4.8;
Matches 87; Conservative 53; Mismatches 153; Indels 135; Gaps 24;

QY 119 GFGGAMTAAALNIALSP-PAONLLKSYFEE-GIGYNIIRVPM-SCDFSIRTY 172
DB 41 GIGAVGGGATSRLLVNTPEYRSQILDYLFKPNFGASLHLKVEIGDGGTDTGTEPSH 100

QY 173 T-YADTPDDFQHLNFSLPEDTKIKIPLIHRALQALQRP-VSLLASPTWSTPLTKTNGAV 230
DB 101 MHYALDENYFRGYEWLWKEAKK-----RNPNTLIGLPWSFFGWL----- 141

QY 231 NKGK----SLKGQPGDIIHQYQWVYVFLDAYAEHLQFVAVTAENEPSAGLLSGYPFQ 286
DB 142 -CKGPDWPNVNLQAYVVTW-----IVGAKRYHDLIDYIGIWNERS----- 184

QY 287 CLGFTPEHORDFIARDLGPTLANSTHNVRLMLDQRLLLPHWAKV---VLTDPAAKY 343
DB 185 -----YNANYI-KILRMLNYQGLQRKIIASDNL-----WBSIASMLLDALEKV 230

QY 344 VHGIHVHYLDLAPAKATLGETHRLFPNTMLFASEACVSKFW-----EOSVRIGS 395
DB 231 VDVIAGH-----YPGTHSAKDAKLTGKWLWSSDFSTLNSDMGAGC 271

QY 396 WDR--GMQYSHSIITNLLHYVGVTDWNLN-----PEG-----GPNWVRNF-VDSP 440
DB 272 WGRILNQNTINGYMTSTI-----AWNLSVSYEQPLPYGRCGLMTAQPWGSVGVVWESP 324

QY 441 IIVDITKDFYKQPMFYHL---GHFSKFIPEGSRVGLVASQKNDLDAVALMHPDGSVAV 497
DB 325 VVWS-AHTQFTQPGWYLYKTVGHLEK---GGSY-----VALTDGLGNLTI 366

QY 498 VVLRSSK 505
DB 367 ILETNSHK 374

RESULT 14
PHK_LACIA STANDARD; PRT; 822 AA.
AC Q9CFH4:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable phosphoketolase (EC 4.1.2.-).
OS L11502.
GN Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
```

```

RL Genome Res. 11:731-753(2001).
CC -!- COFACTOR: Thiamine pyrophosphate (Potential).
CC -!- SIMILARITY: BELONGS TO THE XFP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE006381; AAK05600.1; -.
CC InterPro; IPR000399; TPP_enzyme.
CC PROSITE; PS60002; PHOSPHOKETOLASE.1; 1.
CC PROSITE; PS60003; PHOSPHOKETOLASE.2; 1.
CC PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SQ _SEQUENCE 822 AA; 93363 MW; C686D569D3E8E22F CRC64;

Query Match 3.3%; Score 94.5; DB 1; Length 822;
Best Local Similarity 20.5%; Pred. No. 6.4;
Matches 103; Conservative 61; Mismatches 180; Indels 159; Gaps 27;

QY 102 TGLLLTLQPEQKFKVKGFGAMT-----DAALNILAL-----SPP- 138
DB 156 TGAILD-QPEQIAFAVVGDEAETGLMTSWHSIKFINPKNDGAILPILDNGFKISNPT 214

QY 139 ----AQNLLKSYFEEGIGYNIIRVPMASCDFSIRTYVADTPDDFQHLNFSLPEDTK 194
DB 215 LFARTSDVDLRKFF--EGLGYS-----PRYIENDDIHDMAYH-KLAAEVED 258

QY 195 LKTIPLIHRALQALQRP-----PVSLASP--WTSPTWLKTNGAVNGKSLGQP 240
DB 259 KATEDIHQIKDAREDNRYONGEIPAWPIVIARLPKGGGPRYNDWSGP---KFDGKGM 315

QY 241 GDYHQWARY-----FVKFLDAYAEHLQFVAVTAENEPSAGLLSGYPF 285
DB 316 --IEHSPRAHQVPLPLSKNKGMTLPEFVKWMTSYQPTL-----F 353

QY 286 QCLGFTPEHORDFIARDLGPTLANSTHNVRLMLDQRLLLPHWAKV---LTDPEAAKY 343
DB 354 NADGSLKEELRDFAKCEKMASNPVTNG---GVDSNLVLPDQWQEFANPISNNRGL 409

QY 344 VHGIHVHYLDLAPAKATLGETHRLFPNTW-LFASACVSGKFWEO-SYRLGWS----- 396
DB 410 LPDNDNMNMNLSKYFA--EIVKLNPTFRFLGPDETMSNRFWEMFKVTNRQWQV 466

QY 397 ---DRGMQYSHSIITNLL--YHVVGWTDWNLNLPNPEGP-----NW 432
DB 467 NPDEFISPEGRIIDSQSEHQAEGLWLE-GYTLTGRTGAFASYESFLRVVDSMLTQHF 525

QY 433 VRNFEVDS-----PIIVDITKDFYKQPMFYHLGHFSKFIPEGSRVGL---VASQKNDL 483
DB 526 IRQAADOKWRHSDYSLNVISTSTVFQD---HNGY-----THQDPGMLTHLAEKSD 575

QY 484 DAVALMHPDGSVAVVVLNRSK 506
DB 576 -IROYLPADGNTLLAVEDRAFQD 597

RESULT 15
TYCC_BACBR
ID TYCC_BACBR STANDARD; PRT; 6486 AA.
AC O30409;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III [Includes: ATP-dependent asparagine
DE adenylyase (AsnA) (Asparagine activase); ATP-dependent glutamine
DE adenylyase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE adenylyase (TyrA) (Tyrosine activase); ATP-dependent valine adenylyase
DE (ValA) (Valine activase); ATP-dependent ornithine adenylyase (OrnA)
DE
```


DE (ornithine activase); ATP-dependent leucine adenyase (LeuA) (Leucine
DE activase)).

GN TYCC.

OS Bacillus brevis.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.

OX NCBI_TaxID=193;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 8185; PubMed=9352938;

RX MEDLINE=9601298;

RA Wozniak H.D.; Marchiel M.A.;

RI The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RI nucleotide sequence and biochemical characterization of functional
RI internal adenylation domains.

RL J. Bacteriol. 179:6843-6850(1997).

CC FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
CC VAL, THR, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC PEPTIDE PRODUCT.

CC -1- COFACTOR CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTHETHEINES (BY
CC SIMILARITY).

CC -1- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.

CC -1- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.

CC -1- DOMAIN: CONSISTS OF SIX MODULES, AND HARBOURS A PUTATIVE
CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLATION, THIOATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).

CC -1- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECANPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU) B, C AND D, IN WHICH PHE AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.

CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.

CC -1- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AF004835; AAC45930.1; --
CC HSSP: P14687; 1AMU
CC InterPro: IPR000873; AMP-bind
CC InterPro: IPR001242; Condensatn
CC InterPro: IPR003680; Pantine-attach
CC InterPro: IPR000379; Ser-estrs.site
CC InterPro: IPR001031; thioesterase.
CC Pfam: PF00501; AMP-binding; 6
CC Pfam: PF00550; pp-binding; 6
CC Pfam: PF00668; Condensation; 6
CC Pfam: PF00975; thioesterase; 1
CC PRINTS: PR00154; AMPBINDING
CC PROSITE: PS00012; PHOSPHOPANTHETHEINE; 6.
CC PROSITE: PS00455; AMP BINDING; 6
CC PROSITE: PS00075; ACP DOMAIN; 6
CC KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.

FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 5157 5157 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 6486 AA; 724011 MW; 4354900AF07DF786 CRC64;
Query Match 3.28; Score 92.5; DB 1; Length 6486;
Best Local Similarity 19.08; Pred No. 1.8e+02;
Matches 80; Conservative 54; Mismatches 127; Indels 161; Gaps 19;
QY 173 TYADTDDFLEN--FSIPEDTKIKLIHRALOLA----- 207
DB 4686 TFPVDPDYPLEQAFMLEDSKLLTLQKMSQVAFPEYFYLDFTVDOEETGNLEH 4745
QY 208 -ORPVSLASPWTSTWLKTNKNGAVNGSKLGGQGDYHQTWARYFVKFLDAYAEHKLQF 266
DB 4746 VAQPNVAIIYTSCTCKPKGVV-----LEHRSYANVAFKDEY--H--- 4787
QY 267 WATAEENPSAGLLSGYPCFCIGTPEHQRDFTARDLGPTLANSTHNVRLMLDDQRL 326
DB 4788 -----LDSFPVRL-----QMASTADV-----STGDFARALLTGGQLVI 4822
QY 327 LPHWAKVLTDPAAKYVHGTAHVWLDLFLAPAKATLGTHRLFN----- 372
DB 4823 CPNGVKM--DP--ASLYETIRRHETIFETATPLMPLMHVYVENELDMQKILLIGA 4877
QY 373 -----TMLF-----ASEACVSKWEQ-----SVPLGSDWQWQYSHSI 406
DB 4878 DSCPAEDFTLLAREGQKMRINSYGVTEACIDTSYEEVDVTAIRSGTVPIGKP----- 4932
QY 407 ITNLLYVVGWTDWNLALNDE-----GGPNWVNFVDSPIIVDITKDTFYKPM--- 455
DB 4933 LPNMTYVV---DAHLNLPVGVGELCIGGAGVARGYLNRP---ELTEKEVVPNFPAG 4986
QY 456 --FVHLGHESKFTIPGSO-----RVGLVASQKNDLDAVALMHPDGSV 496
DB 4987 ERLYRTGLAKWRADGNVEFLGRNDHQVIRGVRIELGEIETQLRLDIT-----EAV 5040
QY 497 VV 498
DB 5041 VV 5042

Search completed: February 20, 2003, 11:18:37

Job time : 22 secs

